

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



W CC

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : H01J 49/04, C12Q 1/00, G01N 33/48		A1	(11) International Publication Number: WO 96/36986 (43) International Publication Date: 21 November 1996 (21.11.96)
(21) International Application Number: PCT/US96/07146 (22) International Filing Date: 17 May 1996 (17.05.96) (30) Priority Data: 08/446,055 19 May 1995 (19.05.95) US 08/447,175 19 May 1995 (19.05.95) US (71) Applicant: PERSEPTIVE BIOSYSTEMS, INC. [US/US]; 500 Old Connecticut Path, Framingham, MA 01701 (US). (72) Inventors: PATTERSON, Dale, H.; Apartment J, 38 Bay Ridge Drive, Nashua, NH 03062 (US). TARR, George, E.; 640 Essex Street, S. Hamilton, MA 01982 (US). (74) Agent: TURANO, Thomas, A.; Testa, Hurwitz & Thibault, L.L.P., High Street Tower, 125 High Street, Boston, MA 02110 (US).			(81) Designated States: JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: METHODS AND APPARATUS FOR SEQUENCING POLYMERS WITH A STATISTICAL CERTAINTY USING MASS SPECTROMETRY			
(57) Abstract <p>The method and apparatus disclosed herein are useful for sequencing polymers using mass spectrometry. The methods involve differing ratios of hydrolyzing agent to polymer disposed upon a reaction surface adapted for use with a mass spectrometer. The methods further involve integrating data obtained from mass spectrometry analysis of a plurality of series of hydrolyzed polymer fragments, and provide statistical interpretation paradigms and computer software therefor. The apparatus involves a mass spectrometer sample holder, having hydrolyzing agent disposed thereon, which is useful for adapting any mass spectrometer for polymer sequencing.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

Title of the Application

METHODS AND APPARATUS FOR SEQUENCING POLYMERS WITH A STATISTICAL CERTAINTY USING MASS SPECTROMETRY

Field of the Invention

The present invention relates generally to methods and apparatus for sequencing polymers, especially biopolymers, using mass spectrometry.

Background of the Invention

5 Biochemists frequently depend on reliable and fast determinations of the sequences of biological polymers. For example, sequence information is crucial in the research and development of peptide screens, genetic probes, gene mapping, and drug modeling, as well as for quality control of biological polymers when manufactured for diagnostic and/or therapeutic applications.

10 Various methods are known for sequencing polymers composed of amino acids, carbohydrates and nucleotides. For example, existing methods for peptide sequence determination include the N-terminal chemistry of the Edman degradation, N- and C-terminal enzymatic methods, and C-terminal chemical methods. Existing methods for sequencing oligonucleotides include the Maxam-Gilbert base-specific chemical cleavage method and the
15 enzymatic ladder synthesis with dideoxy base-specific termination method. Each method possesses inherent limitations that preclude it being used exclusively for complete primary structure identification. To date, Edman sequencing and adaptations thereof are the most widely used tools for sequencing certain protein and peptides residue by residue, while the enzymatic synthesis method is preferred for sequencing oligonucleotides.

20 In the case of protein and peptide sequencing, C-terminal sequencing via chemical methods has proven particularly difficult while being only marginally effective, at best. (See, e.g., Spiess, J. (1986) Methods of Protein Characterization: A Practical Handbook (Shively, J.E. ed., Humana Press, N.J.) pp. 363-377; Tsugita et al. (1994) J. Protein Chemistry 13:476-479).

- 2 -

Consequently, the C-terminus remains a region often not analyzed because of lack of a dependable method.

In the case of both peptides and oligonucleotides, an alternate approach to chemical sequencing is enzymatic cleavage sequencing. In the case of oligonucleotides, over 150 different enzymes have been isolated and found suitable for preparing oligonucleotide fragments. In the case of peptides, serine carboxypeptidases have proven popular over the last two decades because they offer a simple approach by which amino acids can be sequentially cleaved residue by residue from the C-terminus of a protein or a peptide. Carboxypeptidase Y (CPY), in particular, is an attractive enzyme because it non-specifically cleaves all residues from the C-terminus, including proline. (See, e.g., Breddam et al. (1987) Carlsburg Res. Commun. 52:55-63.)

Sequencing of peptides by carboxypeptidase digestion has traditionally been performed by a laborious, direct analysis of the released amino acids, residue by residue. Not only is this approach labor-intensive, but it is complicated by amino acid contaminants in the enzyme and protein/peptide solutions, as well as by enzyme autolysis. A further hindrance to any sequencing effort of this type is the absolute requirement for good kinetic information concerning the hydrolysis and liberation of each individual residue by the particular enzyme used.

With the advancement of mass spectrometric techniques capable of high mass analysis such as field desorption (Hong et al. (1983) Biomed. Mass Spectrom. 10:450-457), electrospray (Smith et al. (1993) 4 Techniques Protein Chem. 463-470), and thermospray (Stachowiak et al. (1988) J. Am. Chem. Soc. 110:1758-1765), it is possible to perform direct mass analysis on large biopolymers such as the peptide fragments resulting from CPY digestion in which the sequence order is preserved, circumventing the need for residue by residue amino acid analysis of the liberated amino acids. In this "ladder" sequencing approach, a sequence can be deduced, in the correct order, by calculating the mass differences between adjacent peptide peaks, the measured differences representing the loss of a particular amino acid residue.

More recently, matrix-assisted laser desorption ionization time-of-flight (MALDI-TOF) mass spectrometry also was shown to be suitable for ladder sequence analysis due to its high sensitivity, resolution, and mass accuracy. Chait et al. ((1993) 262 Science 89-92) exploited these assets of MALDI-TOF in the ladder sequencing of N-terminal ladders formed from partial blockage at each step of chemical digestion by the Edman degradation method. This approach

- 3 -

still suffers from the same limitations of traditional Edman chemistry including the complexity of the process, the time-consuming nature of the process, and the lack of C-terminal information. Yet, it confirms the utility of MALDI-TOF for sequencing peptides using the peptide ladder scenario. Other researchers have also illustrated that carboxypeptidase digestion of peptides can be combined with MALDI-TOF to analyze the resulting mixture of truncated peptide. For example, eight consecutive amino acids have been sequenced from the C-terminus of human parathyroid hormone 1-34 fragment (Schar et al. (1991) Chimia 45:123-126). Additionally, carboxypeptidase digestion of peptides has been combined with other mass spectrometry methods such as plasma desorption (Wang et al. (1992) Techniques Protein Chemistry III (ed., R.H. Angeletti; Academic Press, N.Y.) pp. 503-515).

All of the above-described sequencing approaches, however, require preliminary optimization steps which are both tedious and time-consuming. Additionally, such preliminary optimization steps unnecessarily consume reagents as well as samples of polymer, usually available in limited quantities. Furthermore, the above-described sequencing approaches ultimately rely on a single limited number of mass spectrum spectra and single mass-to-charge ratio data points, which can result in a statistically insufficient basis for determining a final polymer sequence.

It is an object of the present invention to provide methods and apparatus for sequencing polymers, particularly biopolymers, using mass spectrometry and time-independent/concentration-dependent hydrolysis of the polymer. More particularly, it is an object of the present invention to provide a method for obtaining sequence information that incorporates a data interpretation strategy based on integrating mass-to-charge ratio data obtained from a plurality of parallel mass spectra. It is another object of the present invention to provide a rapid method for obtaining sequence information by circumventing the time-consuming optimization and method enhancement required by prior art methods. It is a further object of the present invention to provide sequence information using reduced quantities of total polymer by combining the sensitivity of mass spectrometry with elimination of sample loss by closely integrating hydrolysis with mass spectrometry analysis.

- 4 -

Summary of the Invention

Accordingly, one aspect of the present invention is directed to an integrated method for sequencing polymers using information gathered by mass spectrometry, which substantially overcomes the problems encountered in the related art. As broadly described herein, the invention provides a method for obtaining sequence information about a polymer comprising a plurality of monomers of known mass. One skilled in the art first provides a set of fragments, created by the hydrolysis of the polymer, each set differing by one or more monomers. The difference between the mass-to-charge ratio of at least one pair of fragments is determined. One then asserts a mean mass-to-charge ratio which corresponds to the known mass-to-charge ratio of one or more different monomers. The asserted mean is compared with the measured mean to determine if the two values are statistically different with a desired confidence level. If there is a statistical difference, then the asserted mean difference is not assignable to the actual measured difference. In some currently preferred embodiments, additional measurements of the difference between a pair of fragments are taken, to increase the accuracy of the measured mean difference. The steps of such a method are repeated until one has asserted all desired μ s for a single difference between one pair of fragments. The method is repeated for additional pairs of fragments until the desired sequence information is obtained.

The claimed methods are applicable to any polymer, including biopolymers such as DNAs, RNAs, PNAs, proteins, peptides and carbohydrates and modified forms of these polymers. The set of polymer fragments may be created by hydrolysis of the intermonomer bonds of the polymers. With regard to the aforementioned polymer, the instant invention contemplates both naturally-occurring and synthetic moieties characterized by a series of different monomers. In certain embodiments, the polymer also can be modified. Thus, the invention also contemplates the inclusion of a hydrolyzing agent to cause the hydrolysis. Hydrolyzing agents may be enzymatic or an agent other than an enzyme, and any combinations thereof.

In one currently preferred embodiment, the method of obtaining sequence information about a polymer includes providing a set of polymer fragments created by hydrolyzing said polymer, each fragment differing by one or more monomers of known mass; measuring the mass-to-charge ratio difference x between a pair of fragments. Next, one asserts a mean difference μ , which is related to a known mass-to-charge ratio of one or more monomers, and selects a desired

- 5 -

confidence level for μ . The step of measuring the mass-to-charge ratio difference x between a pair of fragments is repeated to obtain a number of measurements n , thereby to determine the statistical mean mass-to-charge ratio difference \bar{x} between the pair of fragments measured. Using the measured mean \bar{x} , one can then determine the standard deviation s of the measured mean mass-to-charge ratio difference \bar{x} previously determined and calculate a test statistic $t_{\text{calculated}}$ with the following algorithm:

$$t_{\text{calculated}} = \frac{|\bar{x} - \mu| \sqrt{n}}{s}$$

One can then repeat the steps of the method until all desired μ s have been asserted for the mass-to-charge ratio difference between a pair of fragments. Sequence information for the polymer is obtained by repeating the steps of the method for additional pairs of fragments.

In another embodiment disclosed herein, the present invention further provides a method of obtaining sequence information about a polymer comprising a series of different monomers which involves: on a reaction surface, providing at least one amount of a hydrolyzing agent which hydrolyzes said polymer and breaks inter-monomer bonds, and a sample of polymer to form differing ratios of agent to polymer; incubating the same for a time sufficient to obtain a plurality of series of hydrolyzed polymer fragments; performing mass spectrometry on a plurality of the series to obtain mass-to-charge ratio data for hydrolyzed polymer fragments contained in the series; and, as described above, integrating data from a plurality of the series to obtain sequence information characteristic of the polymer sample.

The instant invention contemplates certain embodiments involving hydrolyzing agents capable of hydrolyzing a polymer to form sequence-defining ladders, as well as certain other embodiments having hydrolyzing agents capable of forming polymer maps. In yet other embodiments, the instant invention provides for hydrolyzing the polymer with combinations of such agents, as well as enzymatic and non-enzymatic hydrolyzing agents. In certain currently preferred methods, the hydrolyzing agent is disposed on a reaction surface in an array of discrete separate zones. In some embodiments, sets of polymer fragments are sequenced by hydrolyzing the polymer on a reaction surface having one or more different amounts of a hydrolyzing agent. In a most preferred embodiment, a hydrolyzing agent is provided in spatially separate differing

- 6 -

amounts on the reaction surface such that parallel concentration dependent hydrolysis occurs. In another embodiment, the hydrolyzing agent is disposed as a gradient. In yet another embodiment, the agent is disposed on the reaction surface in a constant amount. In other embodiments, polymer is similarly disposed on the reaction surface. In all embodiments, differing agent to polymer ratios are disposed upon the reaction surface and incubated to obtain a plurality of series of hydrolyzed polymer fragments. The various manners in which such differing ratios can be accomplished will be obvious to the skilled practitioner.

For example, a series of concentrations of hydrolyzing agent can be dispersed across a row of the μ L wells of the sample plate of the Voyager™ MALDI-TOF Biospectrometry Workstation, available from PerSeptive Biosystems, Inc. Following passive evaporation, matrix may be added to each well and the sample plate "read" with a MALDI-TOF mass spectrometer. Although time-dependent and concentration-dependent digestions should yield analogous sequence information, it is preferred to use a concentration-dependent approach because it is easily automated, all samples are ready at the same time, and less sample material is lost due to transfer from reaction vessels to the analysis plate. It is therefore preferred to use concentration-dependent on plate hydrolysis, with subsequent analysis on a MALDI mass spec, because it requires only a few pmol of total peptide as a combined result of the sensitivity of MALDI and no sample loss upon moving from digestion to analysis.

When obtaining sequence information by MALDI, a suitable light-absorbent matrix may be added to the polymer fragments at any time prior to measuring the mass-to-charge ratios. For example, matrix may be preloaded onto the reaction surface, or, alternatively, added to the hydrolyzing mixture, prior to, during, or after hydrolysis.

In certain other embodiments, the method provides also combining the agent and polymer with other useful moieties. In one embodiment, moieties which selectively shift the mass of hydrolyzed fragments prior to mass spectrometry analysis are included. In another embodiment, moieties capable of improving ionization of hydrolyzed fragments are included. In yet another embodiment, the method provides for including a light-absorbent matrix. The instant method also contemplates embodiments in which any one or more of the above-described moieties are combined with the agent and polymer prior to mass spectrometry analysis.

- 7 -

Other aspects of the instant invention are related to apparatus and kits for sequencing polymers. The apparatus and kits of the invention in various embodiments include either a mass spectrometer associated with a computer responsive thereto, or a computer associated with a mass spectrometer. In one embodiment the apparatus of the invention includes a mass spectrometer having a means for generating ions, a means for accelerating ions, and a means for determining ions. The mass spectrometer is associated with a computer which is responsive to the mass spectrometer, wherein the computer has the means for performing the methods of the invention.

The apparatus of the invention in yet other embodiments includes a computer readable disc having thereon the information necessary to, in combination with a mass spectrometer, perform the methods of the invention. In other embodiments, the apparatus includes the computer itself, having means for performing the methods of the invention.

More particularly, one embodiment of the apparatus of the instant invention involves a novel form of sample plate or sample holder for a mass spectrometer. The sample plate or sample holder comprises a reaction surface with spacially separate areas having differing ratios of polymer and hydrolyzing agent. After a suitable incubation period during which the hydrolyzing agent hydrolyzes inter-monomer bonds within the polymer in each area, a plurality, typically all, of the areas containing hydrolyzed polymer fragments are ionized, typically serially, in the mass spectrometer and data representative of the mass to charge ratios of these fragments are obtained. One or more of the areas will have ratios of hydrolyzing agent to polymer suitable for more or less optimal generation of useful ladder elements or other polymer fragments. Some areas on the sample holder may have overly hydrolyzed polymer fragments useless for deriving sequence information. Other areas may contain substantially unhydrolyzed polymer. By mass spectrometry analysis of all areas, however, at least some mass to charge ratio data can be obtained from fragments generated in one or more areas. Thus, by integrating the data from different areas, the method of the invention obviates the necessity to empirically prepare samples to ascertain the appropriate ratio of hydrolyzing agent to polymer, as well as optimal reaction time and carefully controlled reaction temperature, heretofore required. Furthermore, different hydrolyzing agents can be used in different series of areas on the sample holder so as to further generate useful hydrolyzed fragments, and the data from these may also be integrated to improve the sequencing process. When data analysis is implemented by a computer program in accordance with the

- 8 -

instant invention, the whole process can be completed minutes after completion of the above-described incubation.

In certain currently preferred embodiments the mass spectrometer sample plate or sample holder has a planar solid surface with at least one amount of a hydrolyzing agent capable of hydrolyzing a polymer disposed thereon. In one embodiment, the hydrolyzing agent is disposed on the reaction surface in a dehydrated form. In another embodiment, the hydrolyzing agent is immobilized on the reaction surface. In yet another embodiment, the hydrolyzing agent is disposed on the reaction surface in the form of a liquid or gel which is resistant to physical dislocation. In still other embodiments, a light-absorbent matrix is disposed on the surface of the sample holder. Additionally, any one or more of such embodiments of the sample holder may further have microreaction vessels on their surface. Certain embodiments of the above-described sample holders are disposable. It is further contemplated that the reaction surface is fabricated from a variety of substrates and assumes a variety of configurations suitable for use with a mass spectrometer. As disclosed herein, all embodiments of the sample plate or sample holder are useful to adapt a mass spectrometry apparatus for sequencing a polymer.

As will be apparent to the skilled artisan, the methods and apparatus for obtaining sequence information in accordance with the instant invention solve problems encountered with conventional polymer sequence methodologies. As described earlier, peptide ladders created using the traditional solution-phase digestion approach, i.e., aliquots of samples are removed at selected time intervals from enzymatic digests, suffer from a number of disadvantages. For example, large amounts of development time, enzyme and peptide are required to obtain significant digestion in a short amount of time while preserving all possible sequence information. For each peptide from which sequence information is to be derived, a time-consuming method development must be performed prior to the actual sequencing analysis since a set of optimum conditions for one peptide is not likely to be useful for another peptide given the composition-dependent hydrolysis rates of various enzymatic agents such as, for example, CPY. As contemplated by the instant invention, an alternative strategy is to perform the digestion on the MALDI sample surface. For example, when conducting on-plate polymer hydrolysis, e.g., exopeptidase digestions, in accordance with the instant method, the overall polymer sequencing effort is superior to the prior art time-dependent digestions in terms of: inherent simplicity of the method and elimination of laborious optimization requirements; reduced loss of sample due to

- 9 -

transfer from reaction vessel to reaction surface; reduced amounts of enzyme and peptide used; and, particularly important for large-scale application, ease of use/automation. Similarly, the mass spectrometry sample plate or sample holder of the instant invention provides advantages heretofore unavailable to the skilled practitioner. For example, certain embodiments minimize reagent handling and greatly facilitate sample processing. The skilled practitioner need only provide a sample of polymer. Virtually all other experimental parameters are pre-optimized.

The foregoing and other objects, features and advantages of the present invention will be made more apparent from the following detailed description. It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory and are intended to provide further explanation of the invention as claimed. The accompanying drawings are included to provide a further understanding of the invention and are incorporated in and constitute a part of this specification, illustrate several embodiments of the invention, and together with the description serve to explain the principles of the invention.

Brief Description of the Drawings

The foregoing and other objects, features and advantages of the present invention, as well as the invention itself, will be more fully understood from the following description of preferred embodiments, when read together with the accompanying drawings, in which:

FIGURE 1 is an exemplary sample plate or sample holder for MALDI analysis. The wells serve as micro-reaction vessels in which on-plate digestions may be performed. The physical dimensions of the plate are 57 x 57 mm and the wells are 2.54 mm in diameter.

FIGURES 2A, 2B and 2C depict several MALDI spectra from a time-dependent CPY digestion of ACTH 7-38 fragment [FRWGKPVGKKRRPVKVYPNGAEDESAEAFPLE] (SEQ. ID. No. 22) at 1 min (2A), 5 min (2B) and 25 min (2C). The nomenclature of the peak labels denotes the peptide populations resulting from the loss of the indicated amino acids. Peaks representing the loss of 19 amino acids from the C-terminus are observed. The symbol * indicates doubly charged ions and # indicates an unidentified peak at $m/z = 2001.0$ and 2744.4 daltons.

FIGURE 3 is a MALDI mass spectrum representing pooled 15 s, 105 s, 6 min and 25 min quenched aliquots from a time-dependent CPY digestion of ACTH 7-38 fragment. All amino acid losses are observed except for those of Glu(28), Asn(25), and Pro(24) which were present as small peaks in the 6 min aliquot and subsequently diluted to undetectable concentrations in this pooled fraction. All conditions are stated in the text

FIGURES 4A and 4B depict various MALDI spectra from on-plate digestions of ACTH 7-38 fragment at various concentrations of Carboxypeptidase Y (CPY): 6.10×10^{-4} U/ μ L (4A); 1.53×10^{-3} U/ μ L (4B). Panels A and B show the spectra obtained from digests using CPY concentrations of 6.10×10^{-4} and 1.53×10^{-3} Units/ μ L, respectively. Laser powers significantly above threshold were used to improve the signal-to-noise ratio of the smaller peaks in the spectrum at the expense of peak resolution. The symbol * indicates doubly charged ions and # indicates an unidentified peak at $m/z = 2517.6$ daltons.

FIGURES 5A, 5B, and 5C depict various MALDI spectra of the following three selected peptides: osteocalcin 7-19 fragment [GAPVPYPDPLEPR] (SEQ. ID. No. 13) (5A), angiotensin 1 [DRVYTHPFHL] (SEQ. ID. No. 8) (5B), and bradykinin [RPPGFSPFR] (SEQ. ID. No. 5) (5C) resulting from on-plate digestions using CPY concentrations of 3.05×10^{-3} ,

- 11 -

3.05×10^{-4} , and 6.10×10^{-4} Units/ μL , respectively. The symbol Na denotes a sodium adduct peak and # denotes a matrix peak at $m/z = 568.5$ daltons.

FIGURES 6A-6E depict various MALDI spectra of exonuclease hydrolysis of a nucleic acid polymer (SEQ. ID. No. 23) at various concentrations of Phosphodiesterase I (Phos
5 I): 0.002 $\mu\text{U}/\mu\text{L}$ (6A); 0.005 $\mu\text{U}/\mu\text{L}$ (6B); 0.01 $\mu\text{U}/\mu\text{L}$ (6C); 0.02 $\mu\text{U}/\mu\text{L}$ (6D); 0.05 $\mu\text{U}/\mu\text{L}$ (6E).

FIGURE 7 depicts a MALDI spectrum of a hydrolyzed nucleic acid polymer (SEQ. ID. No. 23) combined with a light-absorbent matrix.

- 12 -

Detailed Description of Preferred Embodiments

As will be described below in greater detail, the instant invention relates to methods, kits and apparatus for sequencing polymers using mass spectrometry. The present invention provides an integrated strategy for obtaining sequence information about a polymer comprising a plurality of monomers of known mass. Specifically, using sets of polymer fragments and mass spectrometry, the invention provides a method of interpretation of sequence data obtained by mass spectrometry which allows the rapid, automated and cost effective sequencing of polymers with a statistical certainty. The present invention further provides methods which utilize polymers and hydrolyzing agents disposed upon a reaction surface. The hydrolyzing agents are enzymatic or non-enzymatic. The hydrolyzing agents react with the polymer to produce sequence-defining polymer ladders or polymer maps. The methods of this invention further involve the step of obtaining mass spectrometry data relating to hydrolyzed polymer series and integrating the data from a plurality of polymer series to determine the polymer sequence. The mass spectrometry method of this invention is applicable to all manner of ion formation and all modes of mass analysis. The kits and apparatus of this invention relate, in part, to a mass spectrometer sample plate or sample holder for adapting a mass spectrometer to obtain sequence information about a polymer in accordance with the method of the instant invention. Specifically, the sample plate has disposed thereon hydrolyzing agent, in dehydrated, immobilized, liquid and/or gel form, and/or a light-absorbent matrix. Optionally, certain of the sample plates of the instant invention are disposable. Other embodiments of the apparatus of the instant invention relate to mass spectrometers, computers and computer discs suitable for use with the aforementioned methods of sequencing polymers.

As used herein, a "polymer" is intended to mean any moiety comprising a series of different monomers suitable for use in the method of the instant invention. That is, any moiety comprising a series of different monomers whose intermonomer bonds are susceptible to hydrolysis are suitable for use in the method disclosed herein. For example, a peptide is a polymer made up of particular monomers, i.e., amino acids, which can be hydrolyzed by either enzymatic or chemical agents. Similarly, a DNA is a polymer made up of other monomers, i.e., bases nucleotides, which can be hydrolyzed by a variety of agents.

- 13 -

A polymer can be a naturally-occurring moiety as well as a synthetically-produced moiety. In a currently preferred embodiment, the polymer is a biopolymer selected from, but not limited to, the following group: proteins, peptides, DNAs, RNAs, PNAs (peptide nucleic acids), carbohydrates, and modified versions thereof.

5 “Sequence information” as used herein is intended to mean any information relating to the primary arrangement of the series of different monomers within the polymer, or within portions thereof. Sequence information includes information relating to the chemical identity of the different monomers, as well as their particular position within the polymer. Polymers with known primary sequences, as well as polymers with unknown primary sequences, are suitable for use in
10 the methods of the instant invention. It is contemplated that sequence information relating to terminal monomers as well as internal monomers can be obtained using the methods disclosed herein. In certain applications, sequence information can be obtained using a sample of an intact, complete polymer. In other applications, sequence information can be obtained using a sample containing less than the intact complete polymer, for example, polymer fragments. Such
15 fragments can be naturally-occurring, artifacts of isolation and purification, and/or generated in vitro by the skilled artisan. Additionally, polymer fragments can be initially derived from and prepared by a variety of fractionation and separation methods, such as high performance liquid chromatography, prior to use with the methods of the instant invention.

The “reaction surface” of the instant method includes any surface suitable for hydrolyzing
20 the subject polymer with the subject agent. The reaction surface can be fabricated from a variety of substrates, such as but not limited to: metals, foils, plastics, ceramics, and waxes. All reaction surfaces must be suitable for use with a mass spectrometer apparatus. The reaction surface of the instant invention can assume any configuration suitable for use with a particular mass spectrometer apparatus. For example, the reaction surface can be a planar solid surface.
25 Alternatively, the surface may have microreaction vessels disposed thereon. In yet another embodiment, the reaction surface can assume the configuration of a probe suitable for use with certain mass spectrometer apparatus. In some embodiments, the skilled artisan will appreciate that the reaction surface can be activated and/or derivatized to enhance or facilitate polymer sequencing in accordance with the instant invention.

- 14 -

The instant invention relates to a method of data analysis of the mass-to-charge ratios obtained by mass spectrometry. As exemplified below in further detail, the method provides a set of fragments, created by hydrolysis of the polymer, each set differing by one or more monomers. The difference between the mass-to-charge ratio of at least one pair of fragments is determined. One then asserts a mean mass-to-charge ratio which corresponds to the known mass-to-charge ratio of one or more different monomers. The asserted mean is compared with the measured mean to determine if the two values are statistically different with a desired confidence level. If there is a statistical difference, then the asserted mean difference is not assignable to the actual measured difference. In some embodiments, additional measurements of the difference between a pair of fragments are taken, to increase the accuracy of the measured mean difference. The steps of the method are repeated until one has asserted all desired mean differences for a single difference between one pair of fragments.

The above-described method is repeated for additional pairs of fragments and the mass-to-charge ratio data from a plurality of parallel mass spectra are integrated until the desired sequence information is obtained. Thus, in its broadest aspect, the claimed invention is an integrated method for generating sequence information about a polymer comprising a plurality of monomers of known mass. The method involves the interpretation of mass-to-charge ratio data of a set of fragments obtained from the polymer, to statistically identify monomer differences between pairs of fragments. In the past, known molecular masses have been compared to MALDI derived masses for a few mass measurements, and researchers have attempted to make general statements on the instrumental mass accuracy. In general, the methods of the claimed invention involve multiple integrated steps which may be automated according to the invention.

After providing a set of polymer fragments, each differing by one or more monomers, the difference, x , between the mass-to-charge ratio of at least one pair of fragments is measured. Next, one asserts a mean difference μ between the mass-to-charge ratio of the pair of fragments measured, wherein μ corresponds to a known mass-to-charge ratio of one or more differing monomers. One then analyses x to determine if it is statistically different from the μ with a selected confidence level.

If one determines that a statistical difference does exist, then the asserted μ is not assignable to the mass difference x with the selected confidence level.

- 15 -

The steps described above are repeated until all desired μ s have been asserted, and then can be repeated for additional pairs of fragments.

In certain embodiments, the analysis to determine if x is statistically different from μ comprises taking repeated measurements of x , a number of times n , to determine a measured mean mass-to-charge ratio difference x between at least one pair of fragments. A standard deviation s of the measured mean x can then be determined, and the measured mean x compared to the asserted mean μ to determine if they are statistically different with the desired confidence level.

In certain embodiments of the present invention, a set of polymer fragments are obtained, either by on plate digestion, or from an external source, and one or more measurements of the mass-to-charge ratio of a pair of the fragments are taken. Peaks representing the loss of one or more monomers can be analyzed using t-statistics to allow assignments to be made with a desired confidence interval. The two-tailed t-test for one experimental mean,

$$t_{\text{calculated}} = \frac{|\bar{x} - \mu| \sqrt{n}}{s}$$

where \bar{x} is the experimental mean mass difference, μ is the asserted mass difference, N is the number of replicates performed and s is the experimental standard deviation of the mean, is applied. All conceivable masses (single residue, di-residue, tri-residue...etc., as well as modified residue masses) are used as μ , the asserted mass, to generate a list of $t_{\text{calculated}}$ values that are then compared against tabulated values for given confidence intervals. All masses that do not statistically differ from the asserted mass, $t_{\text{calculated}} < t_{\text{table}}$, are statistically assigned to that residue(s) at the given level of confidence. This information can be used to check hypothesized compositions or used to search a database for a sequence. When performing database searching, these levels of confidence can be used in the search algorithm as a tool to aid in obtaining quality "hits."

Ultimately, this technique is to be used for the sequence determination of peptides of unknown sequence. By comparing the known molecular masses to the MALDI derived masses for a few mass measurements, researchers have attempted to make general statements of

- 16 -

instrumental mass accuracy (e.g. better than 0.1%). Ascribing this mass accuracy to any individual mass measurement for the purpose of residue assignment holds no statistical validity, therefore making true residue assignment and direct application to unknowns difficult. In order to call amino acid sequences by ladder sequencing/MALDI strategies, statistical levels of confidence
5 must be placed on residue assignments.

It is contemplated as disclosed herein that the above-described method of integrating data can further comprise the steps of: providing, on a reaction surface, at least one amount of hydrolyzing agent which hydrolyzes a polymer to break intermonomer bonds and produce a set of polymer fragments, and a sample of the polymer such that differing ratios of agent to polymer are
10 formed on the reaction surface; incubating the combined polymer and agent for a time sufficient to obtain a plurality of series of hydrolyzed polymer fragments; and, performing mass spectrometry on a plurality of the series to obtain mass-to-charge ratio data.

For example, a set of polymer fragments created by the endohydrolysis of a polymer can be used to practice the instant invention. Typically, the use of an endohydrolase creates a set of
15 fragments defining a map of said polymer. The mass-to-charge ratio of the fragments is measured, and a hypothetical identity is asserted for the fragment measured. The hypothetical identity corresponds to a known identity of a fragment of a reference polymer. Information on reference polymers is easily included in a database to be used with this method. After selecting a desired confidence level, one determines whether the mass-to-charge ratio of the asserted
20 hypothetical fragment is statistically different from the mass-to-charge ratio of the asserted hypothetical fragment. If it is, then the steps are repeated for different additional hypothetical fragments. This method is repeated until sufficient information is obtained about the fragments that one can identify the polymer with a desired confidence level. Thus, when one is working with maps, one essentially determines whether the fragments of the polymer corresponds to fragments
25 of a known polymer with enough certainty to identify the polymer. It is preferable that the hypothetical identities which are asserted correspond to a known identity derived from a computer database of known sequences.

The methods of the invention also contemplate providing multiple different sets of fragments of the same polymer, i.e. maps and ladders, to obtain the maximum amount of sequence
30 information possible.

- 17 -

The sets of polymer fragments can be created by any method. Certain of the claimed methods contemplate the step of hydrolyzing the polymer with a hydrolyzing agent to obtain the fragments, or synthesizing fragments, as well as merely providing a set of fragments which have been obtained previously. As used herein, the term "hydrolyzing agent" is intended to mean any agent capable of disrupting inter-monomer bonds within a particular polymer. That is, any agent which can interrupt the primary sequence of a polymer is suitable for use in the methods disclosed herein. Hydrolyzing agents can act by liberating monomers at either termini of the polymer, or by breaking internal bonds thereby generating fragments or portions of the subject polymer.

Generally, a preferred hydrolyzing agent interrupts the primary sequence by cleaving before or after a specific monomer(s); that is, the agent specifically interacts with the polymer at a particular monomer or particular sequence of monomers recognized by the agent as the preferred hydrolysis site within the polymer. All of the currently preferred hydrolyzing agents described herein are commercially available from reagent suppliers such as Sigma Chemicals (St. Louis, MO).

In some preferred embodiments, an excipient is added to, and used in conjunction with, the hydrolyzing agent. The excipients contemplated herein facilitate lyophilization and/or dissolution of the hydrolyzing agent. For example, fucose and other sugars suitable for use with the instant invention are contemplated. Suitable for use is intended to mean that no interference with mass spectrometry is encountered by the use thereof. Other excipients useful in the instant invention are pH modifiers, such as ammonium acetate. Still other excipients suitable for use in the methods and apparatus disclosed herein are those which act as stabilizers of the integrity of the hydrolyzing agent. With respect to excipients, the identity of those suitable will be obvious to the skilled artisan using only routine experimentation. While certain preferred excipients are described above, identification of suitable equivalents is within the skill of the ordinary artisan.

In one currently preferred embodiment, the hydrolyzing agent is a hydrolase enzyme. Some hydrolases are endohydrolases, others are exohydrolases. The particular hydrolase used is determined by the nature of the polymer and/or the type of sequence information desired. Its identity can be readily determined by the skilled artisan using no more than routine experimentation. For example, currently preferred endohydrolases include but are not limited to: endonucleases, endopeptidases, endoglycosidases, trypsin, chymotrypsin, endoproteinase Lys-C, endoproteinase Arg-C, and thermolysin. Currently preferred exohydrolases include but are not limited to: exonucleases, exoglycosidases, and exopeptidases. The currently preferred

- 18 -

exonucleases include, but are not limited to: phosphodiesterase types I and II, exonuclease VII, λ -exonuclease, T7 gene 1 exonuclease, exonuclease III, BAL-31, exonuclease I, exonuclease V, exonuclease II, and DNA polymerase III. The currently preferred exoglycosidases include, but are not limited to: α -mannosidase I, α -mannosidase, β -hexosaminidase, β -galactosidase, α -fucosidase I, α -fucosidase II, α -galactosidase, α -neuraminidase, α -glucosidase I and α -glucosidase II. The currently preferred exopeptidases include, but are not limited to: carboxypeptidase Y, carboxypeptidase A, carboxypeptidase B, carboxypeptidase P, aminopeptidase 1, LAP, proline aminodipeptidase, leucine amino peptidase, and cathepsin C.

In certain other embodiments, the hydrolyzing agent is an agent other than an enzyme. For example, such an agent can be a chemical, such as an acid. Currently preferred agents other than an enzyme include but are not limited to: cyanogen bromide, hydrochloric acid, sulfuric acid, and pentafluoropropionic fluorohydride. In some embodiments, hydrolysis can be accomplished using partial acid hydrolysis in accordance with the methods disclosed herein. Again, the identity of a hydrolyzing agent other than an enzyme will be determined by the nature of the polymer and the type of sequence information desired. It is within the skilled practitioner's ability to identify a suitable agent, as well as the circumstances under which such an agent is preferred.

The instant method further provides for use of combinations of the above-described individual hydrolyzing agents. For example, combinations of enzymes can be used in the claimed invention. Combinations of hydrolyzing agents other than enzymes can also be used. Furthermore, combinations of enzymes with agents other than enzymes can also be used in the instant method. Again, the exact combination and the circumstances under which such a combination is appropriate will depend upon the nature of the polymer and the sequence information desired. The skilled practitioner will know when combinations of hydrolyzing agents are suitable for use in the methods disclosed herein.

Numerous examples of hydrolyzing agent/polymer sequence-specific interactions are well known in the art. For example, as described above, currently preferred polymers such as proteins and DNAs specifically interact with proteinases and nucleases, respectively. Certain of the preferred proteinases specifically recognize the C-terminus (carboxypeptidase Y) or the N-terminus (amino peptidase 1) of a protein's amino acid sequence. Certain of the preferred nucleases specifically recognize the 5' or the 3' terminus of a polynucleotide's base sequence.

- 19 -

Some nucleases recognize single-stranded polynucleotides; others recognize double-stranded polynucleotides while still others recognize both.

The claimed invention can be applied to the sequencing of any natural biopolymer such as proteins, peptides, nucleic acids, carbohydrates, etc., as well as synthetic biopolymers such as PNA and phosphotiolated nucleic acids. The ladders could conceivably be created enzymatically using exohydrolases, endohydrolases or the Sanger method and/or chemically by truncation synthesis or failure sequencing. It is preferable to use on-plate digestion and interpretation of peptide ladders created from carboxypeptidase Y, carboxypeptidase P and aminopeptidase I digestions of numerous peptides.

In accordance with the instant invention, exohydrolases generate a series of hydrolyzed fragments comprising a sequence-defining "ladder" of the polymer. That is, these agents generate a series of hydrolyzed fragments, each hydrolyzed fragment within the series being a "ladder element," which collectively comprise a sequence-defining "ladder" of the polymer. Ladder elements represent hydrolyzed fragments from which monomers have been consecutively and/or progressively liberated by the exohydrolase acting at one or the other of the polymer's termini. Accordingly, ladder elements are truncated hydrolyzed polymer fragments, and ladders per se are concatenations of these collective truncated hydrolyzed polymer fragments. In this manner, for example, sequence information relating to the amino acid sequence of a protein can be obtained using carboxypeptidase Y, an agent which acts at the carboxy terminus. By using the methods disclosed herein to generate a series of protein hydrolysates related one to the other by consecutive, repetitive liberation of amino acid residues, the skilled artisan can reconstruct the primary sequence of the intact protein polymer as described in further detail below.

Similarly, hydrolyzing agents other than exohydrolases which also act at one or the other of a polymer's termini generate ladder elements which collectively comprise a series of sequence-defining ladders. For example, the well-known Edman degradation technique and associated reagents can be adapted for use with the methods of the instant invention for this purpose. Thus the above-described subtractive-type sequencing method, through which repetitive removal of successive amino-terminal residues from a protein polymer can occur, can also be accomplished with hydrolyzing agents other than enzymes as disclosed herein.

- 20 -

As previously described, sequence information can also be obtained using hydrolyzing agents which act to disrupt internal inter-monomer bonds. For example, an endohydrolase can generate a series of hydrolyzed fragments useful ultimately in constructing a "map" of the polymer. That is, this agent generates a series of related hydrolyzed fragments which collectively contribute information to a sequence-defining "map" of the polymer. For example, peptide maps can be generated by using trypsin endohydrolysis in tandem with cyanogen bromide endohydrolysis to obtain hydrolyzed fragments with overlapping amino acid sequences. Such overlapping fragments are useful for reconstructing ultimately the entire amino acid sequence of the intact polymer. For example, this combination of hydrolyzing agents generates a useful plurality of series of hydrolyzed fragments because trypsin specifically catalyzes hydrolysis of only those peptide bonds in which the carboxyl group is contributed by either a lysine or an arginine monomer, while cyanogen bromide cleaves only those peptide bonds in which the carbonyl group is contributed by methionine monomers. Thus, by using trypsin and cyanogen bromide hydrolysis in tandem, one can obtain two different series of hydrolyzed "mapping" fragments. These series of mapping fragments are then examined by mass spectrometry to identify specific hydrolysates from the second cyanogen bromide hydrolysis whose amino acid sequences establish continuity with and/or overlaps between the specific hydrolysates from the first hydrolysis with trypsin. Overlapping sequences from the second hydrolysis provide information about the correct order of the hydrolyzed fragments produced by the first trypsin hydrolysis. While these general principles of peptide mapping are well-known in the prior art, utilizing these principles to obtain sequence information by mass spectrometry as disclosed herein has heretofore been unknown in the art.

It will be obvious to the skilled artisan that certain sequencing determinations will be best accomplished using the above-described ladder scenario, while others will be better suited to the mapping scenario. In some situations, a combination of the ladder and mapping sequencing methodologies taught herein will provide optimum sequence information. Using only routine experimentation, the skilled artisan will be able to obtain optimum sequence information using the ladder and/or mapping methods in conjunction with mass spectrometry analysis of a plurality of the series of hydrolyzed polymer fragments.

As contemplated by the instant method, a sample of polymer includes biological fluids containing (or suspected to contain) the polymer of interest. As used herein, a sample of polymer

- 21 -

is also intended to include isolated and purified polymer. Additionally, a sample of polymer can be aqueous or non-aqueous.

Adding a sample of polymer to the reaction surface can be accomplished in a variety of ways. For example, the sample can be introduced as individual aliquots, or the sample can be introduced in a continuous mode such as sample eluting from a preparative or qualitative column. In both cases, the sample can be introduced manually or by automated means.

Upon adding a sample of polymer and hydrolyzing agent to the reaction surface, the instant method provides that differing concentrations of agent or ratios of agent to polymer are formed on said reaction surface. For example, if the polymer sample contains a uniform amount of polymer, then the method contemplates that differing amounts of agent be disposed on the reaction surface. This would produce differing agent to polymer ratios. The differing amounts of agent can be in the form of discrete separate zones to which a constant amount of polymer is added. Alternatively, the differing amounts of agent can be in the form of a non-discrete gradient of agent ranging from low to high amounts of agent, perhaps in the form of strip of appropriate length and width. By introducing a strip of polymer of equal length and width which contains a constant amount of polymer, differing agent to polymer ratios are produced. As contemplated herein, the agent and polymer can assume any configuration and be present in any amount(s); all that is required is that the combination of agent and polymer results in differing ratios of the same disposed on the reaction surface. It will be obvious to the skilled artisan that differing ratios of agent to polymer can also be accomplished by disposing a constant amount of agent on the reaction surface and adding varying amounts of polymer, e.g., a polymer gradient or discrete separate zones of differing amounts of polymer or polymer solution. In the case of a polymer gradient, polymer eluted from a column in the form of a gaussian-distributed gradient is currently preferred.

The instant method further provides for incubating the above-described agent to polymer ratios for a time required to obtain the requisite plurality of series of hydrolyzed polymer fragments. Incubating can proceed under any conditions suitable for hydrolyzing the polymer and for any amount of time required to obtain a plurality of series of hydrolyzed fragments. Generally speaking, the disclosed methods permit sequencing information to be obtained in relatively short time periods, for example, in less than 1 hour. The incubation time, however, can be shortened or

- 22 -

lengthened depending upon the nature of the polymer and/or hydrolyzing agent(s). It will be obvious to one skilled in the art how to identify appropriate incubation times and optimize the same. Incubation reactions can be terminated by evaporation.

As used herein, a "plurality of series" of hydrolyzed polymer fragments is intended to mean that hydrolyzed fragments are produced by at least two different agent:polymer ratios, and that each agent:polymer ratio generates a series of hydrolyzed fragments. For example, if a constant amount of polymer is added to two separate zones of agent containing different amounts of agent, each zone represents one agent:polymer ratio and each zone produces one series of hydrolyzed fragments. When taken together, the two zones are a plurality which collectively contain a plurality of series of hydrolyzed polymer fragments. As disclosed and exemplified herein, the instant methods teach obtaining sequence information by performing mass spectrometry on a plurality of series of hydrolyzed fragments to obtain mass-to-charge ratio data for hydrolyzed polymer fragments contained therein. This contemplates that at least two different agent:polymer ratios be provided and analyzed by mass spectrometry.

The claimed invention may be practiced using any type of mass spectrometry known in the art. Moreover, any manner of ion formation can be adapted for obtaining mass-to-charge ratio data, including but not limited to : matrix-assisted laser desorption ionization, plasma desorption ionization, electrospray ionization, thermospray ionization, and fast atom bombardment ionization. Additionally, any mode of mass analysis is suitable for use with the instant invention including but not limited to: time-of-flight, quadrapole, ion trap, and sector analysis. A currently preferred mass spectrometer instrument is an improved time-of-flight instrument which allows independent control of potential on sample and extraction elements, as described in copending U.S.S.N. 08/446,544 (Atty. Docket No. SYP-111) filed on even date herewith and which is herein incorporated by reference. In certain embodiments, the mass spectrometers used to practice the instant invention include a means to generate ions, a means to accelerate ions, and, a means to detect ions. Any ionization method may be used, for example, desorption, negative ion fast atom bombardment, matrix- assisted laser desorption and electrospray ionization. It is preferable to use matrix-assisted laser desorption mass spectrometry.

It is further contemplated that any of the methods of the instant invention as described herein can further comprise the step of eluting from a liquid chromatography column a sample

- 23 -

comprising a polymer or polymer fragments for which sequence information is to be obtained. In such embodiments, the sample eluted from the column is rendered compatible with a mass spectrometer by contact with a suitable buffer prior to the step of determining mass to charge ratio.

5 The method of the instant invention also provides for including moieties useful in mass spectrometry. For example, a light-absorbent matrix can be introduced at any point prior to performing mass spectrometry analysis by laser desorption. Light-absorbent matrices are particularly useful for analysis of biopolymers. Matrix-assisted laser desorption ionization techniques, as well as various matrices suitable therefor, are well known in the art and have been
10 described, for example, in U.S. 5,288,644 (issued February 22, 1994) and U.S.S.N. 08/156,316 (Atty. Docket No. Vestec-14-2, allowed April 18, 1995), the disclosures of which are herein incorporated by reference.

Other moieties useful in the instant method include those capable of selectively shifting the mass of certain hydrolyzed fragments. These, too, can be added at any point prior to mass
15 spectrometry analysis. Currently preferred mass-shifting moieties include, but are not limited to, those moieties which produce reaction products such as: alkyl, aryl, alkenyl, acyl, thioacyl, oxycarbonyl, carbamyl, thiocarbamyl, sulfonyl, imino, guanyl, ureido, and silyl reaction products. Attachment of such moieties to hydrolyzed polymers is achieved using art-recognized attachment chemistries. The particular moiety best suited to a particular sequence determination will depend
20 upon the nature of the polymer and the hydrolyzed fragments. The skilled artisan will be able to determine which moiety to use, if any.

Another group of moieties suitable for use with the instant method are those which can improve ionization of hydrolyzed fragments. Such moieties can be introduced at any time prior to mass spectrometry analysis. Currently-preferred ionization-improving moieties include, but are
25 not limited to, those moieties which produce reaction products such as: amino, quarternary amino, pyridino, imidino, guanidino, oxonium, and sulfonium reaction products. Preparation and/or use of such moieties are well known in the art.

In another aspect, the instant invention provides a mass spectrometer sample plate or sample holder. As used herein, the terms "sample plate" and "sample holder" are used
30 synonymously. The instant sample plate is useful for adapting any mass spectrometer apparatus

- 24 -

for obtaining sequence information in accordance with the disclosed methods. In one currently preferred embodiment, the sample holder has a planar solid surface on which is disposed hydrolyzing agent. In another currently preferred embodiment, the sample holder has the form of a probe useful in certain mass spectrometer apparatus. In all embodiments of the sample plate or
5 holder, the agent can be in dehydrated, immobilized, liquid and/or gel form. In embodiments having agent in liquid or gel form, the agent is resistant to physical dislocation and is chemically stable for at least about one to two months, thereby facilitating both transport and storage. These considerations are particularly useful for commercial applications involving the sample plate of the present invention. Furthermore, the agent can be disposed in separate discrete zones of differing
10 amounts, or in a non-discrete gradient. Alternatively, the agent can be disposed in a constant amount on the surface of the sample plate. In other embodiments, the sample plate has a light-absorbent matrix disposed on its surface; this can be with or without hydrolyzing agent.

In certain currently preferred embodiments of the instant invention, at least one amount of a dehydrated agent capable of hydrolyzing a polymer is disposed on the planar solid surface of the
15 sample plate. Similarly, at least one amount of an immobilized agent capable of hydrolyzing a polymer can be disposed thereon. In still another preferred embodiment, the sample plate has disposed thereon at least one amount of a hydrolyzing agent in liquid or gel form, said liquid or gel form being resistant to physical dislocation.

The sample plate can also have microreaction vessels arranged on its surface. In one
20 embodiment, these vessels can be depressions on the plate's surface resulting from chemical-etching or similar techniques. The sample plate can be fabricated from a variety of substrates including but not limited to: metals, foils, plastics, ceramics, and waxes. In certain embodiments, the sample plate is disposable. In certain other embodiments, the sample plate disclosed herein is a component of a kit useful for sequencing polymers by mass spectrometry.

25 With respect to any of the sample plates or sample holders contemplated herein, the surface can comprise an array of discrete separate zones of differing amounts of said agent. Alternatively, the surface comprises a non-discrete gradient of said agent or a constant amount of said agent.

30 Additionally, any embodiment can further comprise a light-absorbent matrix, and/or microreaction vessels, and/or be fabricated of a disposable material.

- 25 -

In yet another aspect, the instant invention provides a kit having a sample plate or holder comprising a reaction surface, said surface providing differing amounts of a hydrolyzing agent to hydrolyze said polymer into said fragments. In one embodiment, the kit contains a sample plate or holder further comprising a matrix suitable for matrix-assisted laser desorption mass spectrometry.

The claimed invention also relates to other mass spectrometer apparatus and kits for performing the methods above. In one embodiment the apparatus of the invention for obtaining sequence information about a polymer comprises a mass spectrometer having a means for generation ions from a sample, a means for acceleration of ions generated, and a detection means. These basic components are available in numerous embodiments, and therefore, the invention is not limited to a particular type of mass spectrometer. The apparatus additionally comprises a computer responsive to the mass spectrometer comprising a means for determining the mass to charge ratio difference x between a pair of polymer fragments; a means for asserting a mean difference μ between the mass-to-charge ratio of the pair of fragments, wherein μ corresponds to a known mass-to-charge ratio of one or more monomers; and a means for analyzing x to determine if it is statistically different from μ with the desired confidence level, and a means for determining when the desired number of possible μ s have been asserted.

Additionally, the information necessary for the claimed methods can be incorporated onto a computer-readable disc, which can render a computer responsive to a mass spectrometer for performing the analysis. Claimed software will automate the process of acquiring and interpreting the data in an intelligent fashion using software feedback control. The data interpretation software would control the number of acquisitions (minimum of 2) that are required to statistically differentiate multiple candidates for an amino acid assignment. The operator would have control of specifying to what minimum statistical level of confidence the assignment(s) must meet.

Practice of the invention will be still more fully understood from the following examples, which are presented herein for illustration only and should not be construed as limiting the invention in any way.

- 26 -

EXAMPLE 1. MATERIALS AND METHODS**(a) Solution-Phase Digestion of ACTH 7-38 Fragment**

For the time course digestion, 500 pmol of synthetic human adrenocorticotrophic hormone (ACTH) fragment (7-38) [FRWGKPVGKKRRPVKVYPNGAEDESAEAFPLE] (SEQ. ID. No. 22) from Sigma Chemical Company (St. Louis, MO), previously dried down in a 0.5 mL eppendorf vial, was resuspended with 33.3 μ L of HPLC grade water (J.T. Baker, Phillipsburg, NJ). In a previously dried down 0.5 mL eppendorf tube, 3.05 units (one unit hydrolyzes 1.0 μ mol N-CBZ-phe-ala to N-CBZ-phenylalanine + alanine per minute at pH = 6.75 and 25°C) of carboxypeptidase Y from bakers yeast (E.C. 3.416.1), purchased from Sigma, was resuspended with 610 μ L of HPLC grade water. To 20 μ L of the ACTH 7-38 fragment solution was added 10 μ L of the CPY solution to initiate the reaction. The final concentrations were 10 pmol/ μ L ACTH and 1.67×10^{-3} units/ μ L CPY yielding an enzyme-to-substrate ratio of 1.67×10^8 units CPY/mol ACTH (1:37 molar ratio assuming CPY MW = 61,000). Aliquots of 1 μ L were taken from the reaction vial at reaction times of 15 s, 60 s, 75 s, 105 s, 2 min, 135 s, 4 min, 5 min, 6 min, 7 min, 8 min, 9 min, 10 min, 15 min and 25 min. At 25 min, 15 μ L of 5×10^{-3} units/ μ L CPY was added to the reaction vial. Aliquots of 2 μ L were removed at total reaction times of 1 hr and 24 hr. The reaction proceeded at room temperature until 2 min when the temperature was elevated to 37°C. All aliquots were added to 9 μ L of the MALDI matrix, α -cyano-4-hydroxy cinnamic acid (CHCA) from Sigma, at a concentration of 5 mg/mL in 1:1 acetonitrile (ACN):0.1% trifluoroacetic acid (TFA) with the exception of the 1 hr and 24 hr aliquots were added to 8 μ L of the matrix. The final total peptide concentrations of the ACTH digestion aliquots in the matrix solutions were 1 pmol/ μ L. A pooled peptide solution was prepared by combining 2 μ L of the 15 s, 105 s, 6 min and 25 min aliquots. Into individual μ L wells on the MALDI sample plate, 1 μ L of each aliquot solution was placed and allowed to evaporate to dryness before insertion into the mass spectrometer.

(b) On-Plate Digestions:

All on-plate digestions were performed by pipetting 0.5 μ L of the peptide at a concentration of 1 pmol/ μ L into each of ten 1 μ L wells across one row of a sample plate configured similarly to the sample plate manufactured and supplied by PerSeptive BioSystems,

- 27 -

Inc. of Framingham, MA and adapted for use with their trademarked mass spectrometry apparatus known as Voyager™. All peptides listed in Table 1 were purchased from Sigma and were of the highest purity offered. To initiate the reaction in the first well, 0.5 μ L of 0.0122 units/ μ L CPY was added. To the subsequent 9 wells was added CPY at concentrations of 6.10×10^{-3} , 3.05×10^{-3} , 1.53×10^{-3} , 6.10×10^{-4} , 3.05×10^{-4} , 1.53×10^{-4} , 7.63×10^{-5} , 3.81×10^{-5} and 0 units/ μ L, respectively. Mixing was assured in each well by pulling the 1 μ L reaction back and forth through the pipet tip. The reaction was allowed to proceed at room temperature until the 1 μ L total volume evaporated on the plate (approximately 10 min). At such time, 1 μ L of 5mg/mL CHCA in 1:1 ACN:0.1% TFA was added to each well, with no further mixing, and allowed to evaporate for approximately 10 min before mass analysis.

(c) MALDI-TOF Mass Spectrometry:

MALDI-TOF mass analysis was performed using the Voyager™ Biospectrometry™ Workstation (PerSeptive Biosystems, Cambridge, MA). A 28.125 KV potential gradient was applied across the source containing the sample plate and an ion optic accelerator plate in order to introduce the positively charged ions to the 1.2 m linear flight tube for mass analysis. For the data acquisition of the ACTH 7-38 fragment and glucagon digests, a low mass gate was used to prevent the matrix ions from striking the detector plate. For the application of the low mass gate, the guide wire was pulsed for a brief period deflecting the low mass ions (approximately <1000 daltons). All other spectra were recorded with the low mass gate off. To enhance the signal-to-noise ratio, 64-128 single shots from the nitrogen laser (337 nm) were averaged for each mass spectrum. The data presented herein were smoothed using an 11 point Savitsky-Golay second order filter. All data was calibrated using an external calibration standard mixture of bradykinin ($MH^+ = 1061.2$) and insulin B-chain, oxidized ($MH^+ = 3496.9$)(both purchased from Sigma) at concentrations of 1 pmol/ μ L in the 5 mg/mL CHCA matrix solution.

- 28 -

(d) Statistical Mass Assignments:

As described in further detail below, the statistical protocol disclosed herein uses the equation for the two-tailed t-test:

$$t_{\text{calculated}} = \frac{|\bar{x} - \mu| \sqrt{n}}{s}$$

5 where \bar{x} is the average experimental mean, μ is the asserted mean, n is the number of replicates and s is the experimental standard deviation. For the assignment of residues to experimentally derived Δ masses, a $t_{\text{calculated}}$ for each asserted mean mass (each possible amino acid assignment) was compared to the tabulated value for a given confidence interval. A $t_{\text{calculated}} > t_{\text{table}}$ indicated that the experimental mass came from a population possessing a different mean than the asserted
10 mass at the given confidence level.

EXAMPLE 2. SEQUENCING OF BIOPOLYMERS**(a) Solution-Phase Sequencing:**

Figure 2 illustrates the MALDI spectra of the 1 min, 5 min and 25 min time aliquots that were removed from a solution-phase time-dependent CPY digestion of ACTH 7-38 fragment.
15 The nomenclature of the peak labels denotes the peptide populations resulting from the loss of the indicated amino acids. Peaks representing the loss of 19 amino acids from the C-terminus are observed. The symbol * indicates doubly charged ions and # indicates an unidentified peak at m/z = 2001.0 and 2744.4 daltons.

The lack of phase control of the enzymatic digestion creates the peptide ladders that are
20 observed in this figure. After 1 min of digestion (Figure 2A), 9 detectable peptide populations exist including the intact ACTH 7-38 fragment and peptides representing the loss of the first 8 amino acids from the C-terminus. The 5 min aliquot (Figure 2B) shows that the peptide populations representing the loss of Ala(32) and Ser(31) have become much more predominant than the 1 min aliquot. Amino acid losses of 11 residues, Ala(32) through Val(22), are present at
25 this digestion time. Figure 2C shows the final detected amino acids of Lys(21) and Val(20) as 4 major peptide populations are detected. Upon increasing the enzyme concentration 2-fold at 25 min, no further digestion was observed through 24 h. The digestion proceeded through the

- 29 -

Val(20) and stopped at the amino acid run of *peptide-KKRRP*. Although CPY may proceed rapidly through proline (e.g., Pro(24)), the basic residue, arginine, at the penultimate position in this case proved to be a combination refractory to CPY.

The lack of phase control coupled with the varied rates of hydrolysis poses problems
5 unique to enzymatic sequencing. Varying ion intensities for the peaks in Figure 2 are due primarily to the rates of hydrolysis that vary according to the amino acids at the C-terminus and penultimate position. When a residue is hydrolyzed at a low rate compared to the neighboring residues, the concentration and, therefore, signal of the peptide population representing the loss of that residue will be small relative to that of the preceding amino acid. This is seen in the mass
10 spectra given in Figure 2. The cleavage of Ala(34) is shown to be slow resulting in the large signal representing the loss of Phe(35). The hydrolysis of glycine and valine are also shown to be slow as the peaks representing the loss of Ala(27) and Tyr(23) are comparatively more intense than those of Gly(26) and Val(22), respectively.

The prior-art time-dependent method presented herein is the result of extensive method
15 optimization and is optimized for obtaining the maximum sequence information in the shortest amount of time. For this particular optimized case, detectable amounts of all populations were observed over 25 min in the three selected time aliquots. This was not the case for numerous preliminary solution-phase digestions that were performed during the method optimization that led to the choice of these optimized conditions. At higher concentrations of CPY the peaks
20 representing the loss of Glu(28) and Pro(24) were often not observed, indicating that CPY cleaves these residues very readily when alanine and tyrosine are at the penultimate positions, respectively. Lower concentrations of CPY allowed for all amino acids to be sequenced but often required long periods of time, e.g., days, for sufficient digestion. In the instance disclosed herein, an enzyme-to-substrate ratio of 1.67×10^8 units CPY/mole peptide was finally found to offer
25 sufficient sequence information in 25 min of digestion.

Alternatively, upon pooling aliquots from 15 s, 105 s, 6 min, and 25 min of total reaction time, MALDI analysis shows that a peptide ladder is formed that contains peaks that represent the loss of almost all amino acids from the C-terminus (Figure 3). All amino acid losses are observed except for those of Glu(28), Asn(25), and Pro(24) which were present as small peaks in the 6 min
30 aliquot and subsequently diluted to undetectable concentrations in this pooled fraction.

- 30 -

A sequence gap is observed here as the peptide populations representing the loss of Glu(28), Asn(25) and Pro(24) exist below a signal-to-noise ratio of 3. These populations were observed as small peaks in the 6 min aliquot mass spectrum but, upon the 4-fold dilution with the other aliquots, exist in too small a concentration to be detected. This emphasizes the necessity of recording individual mass spectra for each time aliquot. The less time-demanding procedure of recording a single spectrum representing pooled results not only created sequence gaps, but lost the time-dependent history of the digestion.

As illustrated above, solution-phase digestion suffers from a number of disadvantages. A large amount of time, enzyme and peptide is required for method optimization in order to obtain significant digestion in a short amount of time while preserving all possible sequence information. For each peptide from which sequence information is to be derived, some time-consuming method development must be performed since a set of optimum conditions for one peptide is not likely to be useful for another peptide given the composition-dependent hydrolysis rates of CPY. An alternative strategy is to perform the concentration-dependent hydrolysis on the MALDI sample surface as described below.

(b) On-Plate Sequencing:

Figure 1 depicts a Voyager™ sample plate for MALDI analysis comprised of a 10 x 10 grid of 1 μ L wells etched into the stainless steel base. These wells serve as micro-reaction vessels in which on-plate digestions may be performed. The physical dimensions of the plate are 57 x 57 mm and the wells are 2.54 mm in diameter.

Half- μ L amounts of both enzyme and substrate were placed in a well and mixed with the pipet tip. The digestion continued for about 10 min until solvent evaporation terminates the reaction. At this time, the digestion mixture was resuspended by placing 1 μ L of the matrix in the well. Since the CHCA matrix is solubilized in 1:1 ACN:0.1% TFA, both hydrophilic and hydrophobic peptide populations from the digest mixture should be resuspended with the low pH prohibiting any further CPY activity. The matrix crystal formation does not appear to be altered (as compared to the time-course experiment) by performing the digestion on-plate. This on-plate strategy significantly decreased the method optimization time by allowing multiple concentration-dependent (time-dependent) digestions to be performed in parallel. Also, sample losses upon

- 31 -

transfer(s) from reaction vial to analysis plate were circumvented using the on-plate approach as all digested material is available for mass measurement.

MALDI spectra corresponding to the on-plate concentration dependent digestions of the ACTH 7-38 fragment for CPY concentrations of 6.10×10^{-4} , and 1.53×10^{-3} units/ μ L, respectively, are illustrated in panels A and B of Figure 4. Panel A and B show the spectra obtained from digests using CPY concentrations of 6.10×10^{-4} and 1.53×10^{-3} units/ μ L, respectively. Laser powers significantly above threshold were used to improve the signal-to-noise ratio of the smaller peaks in the spectrum at the expense of peak resolution. The symbol * indicates doubly charged ions and # indicates an unidentified peak at m/z - 2517.6 daltons.

The lower concentration digestion yielded 12 significant peaks representing the loss of 11 amino acids from the C-terminus. The digestion from the higher concentration of CPY showed some overlap of the peptide populations present at the lower concentration as well as peptide populations representing the loss of amino acids through the Val(20). The concentration of the peptides representing the loss of the first few amino acids have decreased to undetectable levels (approximately <10 fmol) with the exception of the Leu(37) peak. By integrating the information in both panels, the ACTH 7-38 fragment sequence can be read 19 amino acids from the C-terminus without gaps, stopping at the same amino acid run of *peptide-RRKKP* as the time-dependent digestion. Figure 4 represents 2 of the 9 CPY concentrations that were performed simultaneously. The method optimization, in this case, was inherent in the strategy. The total time of method development (optimal digestion conditions), digestion, data collection and data analysis was under 30 min using this on-plate approach. The consumption of both peptide and enzyme was minimal as a total of 5 pmol of total peptide was digested across the 10 well row containing 9 digestions and 1 well with peptide plus water. Also, only 1.97 pmol of CPY (assuming 100 unit/mg and MW = 61,000) was required for the entire experiment.

- 32 -

Table 1

Peptide	SEQ ID Nos.	Sequence	Average Mass ¹	Charge ²	Polarity
Sleep Inducing Peptide	1	WAGGDASGE	848.8	-2.0	polar
Amino Terminal Region of Hbs β chain ³	2	VHLTPVEK	922.1	+0.5	mid
Interleukin-1 β 163-171 Fragment ³	3	VQGEESNDK	1005.0	-2.0	polar
TRH Precursor	4	KRQHGPGR	1006.2	+4.5	very
Bradykinin	5	RPPGFSPYR	1061.2	+2.0	mid
Lutenizing Hormone Releasing Hormone ³	6	pyro.EHWSYGLRPG.amide	1182.3	+1.5	mid
Physalaemin	7	pyro.EADPNKFYGLM.amide	1265.4	0	mid
Angiotensin I	8	DRVYIHPFHL	1295.5	+1.0	non
Renin Inhibitor	9	PHPFHFFVYK	1318.5	+2.0	non
Kassinin	10	DVPKSDQFVGLM.amide	1334.5	-2.0	non
Substance P	11	RPKPQQFFGLM.amide	1347.6	+3.0	mid
T-Antigen Homolog	12	CGYGPKKKRKVGG	1377.7	+5.0	polar
Osteocalcin 7-19 Fragment	13	GAPVPYPDPLEPR	1407.6	-1.0	mid
Fibrinopeptide A	14	ADSGEGDFLAEGGGVR	1536.6	-3.0	mid
Thymopoietin II 29-41 Fragment	15	GEQRKDVPVQLYL	1610.8	0	polar
Bombesin	16	pyro.EQRLGNQW(AVGH)LM.amide	1619.9	+1.5	mid
ACTH 11-24 Fragment	17	KPVGKKRRPVKVYP	1652.1	+6.0	mid
α -Melanocyte Stimulating Hormone	18	acetyl.STSMEHFRWGKPV. amide	1664.9	+1.5	mid
Angiotensinogen 1-14 Fragment	19	DRVYIHPFHLVYS	1759.0	+1.0	non
Angiogenin	20	ENGLPVHLDQSI(FR)R	1781.0	+0.5	mid
Glucagon	21	HSQ...DSRRRAQDFVQW(LMN)T	3482.8	+1.0	polar
ACTH7-38 Fragment	22	FRW...RRPVKVYPNGAEDESAAEF PLE	3659.15	+2.0	polar

1 calculated

2 at pH 6.5

3 no sequence information was obtained

- 5 Listed in Table 1 are the peptides that have been digested and analyzed using this novel on-plate strategy. These peptides were selected to represent peptides of varying amino acid composition, size (up to MW = 3659.15), charge and polarity. The bolded amino acids indicate that a peak representing the loss of that residue was observed in one or more of the MALDI spectra taken across the row of digestions. In order to be able to identify a residue, the peak
- 10 representing the loss of that amino acid and the preceding amino acid must be present. The residues that are enclosed in parenthesis are those for which the sequence order could not be deduced. Overall, CPY offered some sequence information from the C-terminus for most of the peptides digested, lending no sequence information in only three of the 22 cases. In two of these three cases, the C-terminus was a lysine followed by an acidic residue at the penultimate position.
- 15 CPY has been reported to possess reduced activity towards basic residues at the C-terminus, and the presence of the neighboring acidic residue seems to further reduce its activity. In the case of the lutenizing hormone releasing hormone (LH-RH), the C-terminal amidated glycine followed by proline at the penultimate position inhibited CPY activity which agrees with reports of CPY slowing at both proline and glycine residues (Hayashi et al. (1975) *J. Biochem.* 77:69-79;

- 33 -

Hayashi, R. (1976) Methods Enzymol. 45:568-587). CPY is known to hydrolyze amidated C-terminal residues of dipeptides and is shown here to cleave those of physalaemin, kassinin, substance P, bombesin, and α -MSH.

As illustrated by the data in Table 1, CPY was able to derive sequence information from all of the peptides, except LH-RH, that possess blocked N-terminal residues (physalaemin, bombesin and α -MSH). This is significant as these peptides would lend no information to the Edman approach. A number of the peptides were sequenced until the detection of the truncated peptide peaks were impaired by the presence of CHCA matrix ions (<600 daltons). The sequencing of the other peptides did not go as far as a combination of residues at the C-terminus and penultimate position that inhibited CPY activity were encountered. Bombesin, angiogenin and glucagon gave gaps in the sequence as residues that were cleaved slowly were followed by residues hydrolyzed more rapidly, as discussed above. The feasibility of the on-plate CPY digestion/MALDI detection strategy appeared to be independent of the overall polarity and charge of the peptide.

Figure 5 shows selected on-plate digestions of osteocalcin 7-19 fragment, angiotensin 1 and bradykinin resulting from on-plate digestions using CPY concentrations of 3.05×10^{-3} , 3.05×10^{-4} , and 6.10×10^{-4} units/ μ L, respectively. The symbol Na denotes a sodium adduct peak and # denotes a matrix peak at m/z - 568.5 daltons.

Each spectrum represents the results of one of the 9 digestions that was performed across the row of wells. In the case of the osteocalcin 7-19 fragment, CPY can proceed through proline (Martin, B. (1977) Carlsburg Res. Commun. 42:99-102; Breddam et al. (1987) Carlsburg Res. Commun. 52:55-63; Breddam, K. (1986), Carlsburg Res. Commun. 51:83-128; Hayashi, R. (1977) Methods Enzymol. 74:84-94; Hayashi et al. (1973) J. Biolog. Chem. 248:2296-2302); the presence of Asp and His at the respective penultimate positions of the two peptides prohibited further CPY activity. Bradykinin is shown to sequence until the matrix begins to interfere with peak detection. For all three of the selected peptides, the total sequence information obtained for the overall 9 well digestion is represented in the single digestion shown. For many other peptides this was not the case. The total sequence information is often derived from 2 or more of the wells as is the case with ACTH 7-38 fragment given in Figure 4.

EXAMPLE 3. STATISTICAL ANALYSIS OF LADDER SEQUENCING BY MALDI**(a) General Principles of Statistical Analysis According to the Instant Invention**

As disclosed above, once the truncated ladders have been formed, matrix is added to the well and multiple measurements were taken from the wells in which peaks representing the loss of an amino acid(s) are present. Statistical interpretation involving the use of t-statistics then allowed assignments to be made with an associated confidence interval. The two-tailed test for one experimental mean,

$$t_{\text{calculated}} = \frac{|\bar{x} - \mu| \sqrt{n}}{s}$$

where \bar{x} is the experimental mean mass difference, μ is the asserted mass difference, n is the number of replicates performed, and s is the experimental standard deviation of the mean, was applied. All conceivable masses (single residue, di-residue, tri-residue, etc., as well as modified residue masses) were used as μ , the asserted mass, to generate a list of $t_{\text{calculated}}$ values that were then compared against tabulated values for given confidence intervals. All masses that did not statistically differ from the asserted mass, $t_{\text{calculated}} < t_{\text{table}}$, were statistically assigned to that residue(s) at the given level of confidence. This information was used to check hypothesized composition or used to search a database for a sequence. When performing database searching, these levels of confidence can be used in the search algorithm as a tool to aid in obtaining quality "hits."

Additionally, the interpretation of data utilized an automated process of acquiring and interpreting the data using software feedback control. The data interpretation software controls the number of acquisitions (minimum of 2) that are required to statistically differentiate multiple candidates for an amino acid assignment. The operator has control of specifying to what minimum statistical level of confidence the assignment(s) should meet.

(b) Analysis of Experimentally-Obtained Mass-to-Charge Ratio Data: Peptides

The use of MALDI for the analysis of truncated ladders as disclosed herein is critical for obtaining accurate sequence data. In the prior art, the technique has been used almost exclusively

- 35 -

to sequence peptides of a defined sequence for which the mass accuracy of the measurement is of little importance. In contrast, the methods disclosed herein are useful for the sequence determination of peptides of unknown sequence. By comparing known molecular masses to the MALDI derived masses for only a few mass measurements, artisans previously have made only
5 general statements of instrumental mass accuracy (e.g., better than 0.1%), but, ascribing this mass accuracy to any individual mass measurement for the purpose of residue assignment holds no statistical validity. Therefore, true residue assignment and direct application to unknowns has heretofore been both difficult and tentative. In order to derive amino acid sequences by ladder sequencing/MALDI strategies, statistical levels of confidence must be placed on residue
10 assignments as disclosed herein.

To place confidence levels on residue assignments, the nature of the experimental errors first must be defined. For systems in which the errors are random, simple t-statistics can be used for amino acid assignment.

To assess the nature of the errors that dominate MALDI analysis of the above-described
15 truncated peptide ladders, the Δ mass differences (i.e., experimental mass difference - actual amino acid mass) for all amino acid assignments made in the 15 aliquots (one spectrum per aliquot) removed from the time-dependent digestion of ACTH 7-38 fragment described above were measured to yield a gaussian distribution with a mean of 0.0089 ± 0.605 ($n=107$). For this
experiment $t_{\text{calculated}} (0.152) < t_{\text{table}} (1.99)$ indicating that the null hypothesis that the average Δ
20 mass difference = 0 cannot be rejected at a 95% confidence level. This indicates that the error is random with no statistically significant systematic error. This is expected as any systematic errors that are present in the mass assignment of individual peptide peaks such as incorrect y-intercept values for two-point mass calibration should cancel out when calculating the mass difference of two adjacent peaks. There are possible systematic components of error that would not be
25 canceled such as incorrect computation of the mass center of one of a set of two adjacent peaks due to partial resolution of the isotopes. This phenomenon was circumvented by the use of a smoothing filter such that all peaks were detected at the actual average mass values.

- 36 -

Table 2

Amino Acid (position)	Actual Mass ¹	Experimental Mass ^{1,2}	Replicates
val(20)	99.13	98.97 ± 0.52 (1.29)	3
lys(21)	128.17	128.15 ± 0.48 (0.44)	7
val(22)	99.13	99.20 ± 0.35 (0.27)	9
tyr(23)	163.17	162.43 ± 0.11 (0.99)	2
pro(24)	97.12	97.49 ± 0.14 (1.25)	2
asn(25)	114.10	114.21 ± 0.82 (0.69)	8
gly(26)	57.05	57.22 ± 0.88 (0.68)	9
ala(27)	71.07	70.19 ± 0.49 (4.40)	2
glu(28)	129.12	130.22 ± 0.47 (4.22)	2
asp(29)	115.09	114.81 ± 0.58 (0.41)	10
glu(30)	129.12	129.27 ± 0.61 (0.39)	12
ser(31)	87.08	87.14 ± 0.47 (0.30)	12
ala(32)	71.07	80.94 ± 0.49 (0.51)	6
glu(33)	129.12	129.39 ± 0.42 (0.44)	6
ala(34)	71.07	71.09 ± 0.30 (0.28)	7
phe(35)	147.18	147.03 ± .73 (0.77)	6
pro(36)	97.12	96.83 ± 0.64 (1.18)	4
leu(37)	113.16	113.63 ± 0.54 (1.34)	3
glu(38)	129.12	128.40 ± 0.52 (1.29)	3

1 the masses given are average masses and in units of daltons

2 the uncertainties of the experimental mass measurements are given as standard deviations (those in the parenthesis are 95% confidence intervals of the mean)

5 Table 2 represents a comparison of the actual average masses of the sequenced residues of the ACTH 7-38 fragment and the experimental mass differences with associated standard deviations and 95% confidence intervals calculated for the time-dependent digestion. The number of replicates indicate the number of spectra that possessed the detectable adjacent peaks required for the mass difference measurement of that particular residue. The need for a significant number
10 of measurements in order to estimate the mean is obvious from the table as the 95% confidence level decreases as the square root of the number of measurements. For all of the residues sequenced, the actual mass fell within $\pm 3\sigma$ the experimental mass distribution. Calculated t-values for each case were less than the tabulated t-value for the 95% confidence interval signifying that the experimental mass is not significantly different than the actual known mass. In
15 order to statistically assign the residues, a calculated t-value for each possible amino acid must be compared with the tabulated value. In other words, the actual masses of all possible amino acids

- 37 -

must be used as an asserted mean, μ , and each null hypothesis (i.e., $\bar{x} - \mu = 0$) made such that a calculated t-value for each possible assignment can be compared to the tabulated value.

Assuming that only the 20 common unmodified amino acids are possible, this was done for the prior art time-dependent ACTH 7-38 fragment digestion. A summary of the results is given in Table 3. The bolded values are those which the experimental mean did not significantly differ from the asserted amino acid mean. Again, the need for adequate population sampling is apparent. There were only two measurements observed for the Glu(28) thereby resulting in a 95% confidence interval of 4.22 daltons (Table 2). This translates into an inability to distinguish between Gln, Lys, Glu and Met (Table 3). The 12 trials that were observed for Glu(30) gave a 95% confidence interval of 0.39 daltons, thereby rendering the Gln, Lys and Met statistically improbable amino acid assignments.

Table 3 represents calculated t-values for 19 sequenced amino acid experimental means in the ACTH 7-38 fragment given the asserted means of 20 common unmodified amino acids. The t_{table} value is given at the end of each column. A $t_{calculated} < t_{table}$ indicates that the experimental mean is not significantly different that the mean of the asserted amino acid at 95% confidence interval. Each $t_{calculated}$ for which this is the case is indicated in bold.

- 38 -

Table 3
ACTH 7-38 Fragment Amino Acid Position

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
Gly							0.58	37.9					69.4		123				
Ala							47.2	2.54				118	0.65		0.18				
Ser					105			48.7				0.44	80.7		141		30.5		
Pro	6.16		17.8		3.74							73.6					0.91		
Val	0.53		0.60		16.6												7.19		
Thr	7.09		16.3																
Cys																		33.6	
Leu/						3.62												1.51	
Ile																			
Asn						0.38			3.87									1.51	
Asp	72.0				3.04			45.5	1.53								4.68	44.3	
Gln	0.11							6.29	72.6									0.90	
Lys	0.11							6.17		6.25				7.12				0.77	
Glu	5.35							3.31	0.85					1.57				2.40	
Met								2.95		11.0				10.6				9.33	
His								20.8								33.2			
Phe																0.50			
Arg				80.4												30.7			
Tyr				9.64															
Trp				305															
<i>t_{table}</i>	4.30	2.45	2.31	12.7	12.7	2.37	2.31	12.7	12.7	2.26	2.20	2.20	2.57	2.57	2.45	2.57	3.18	4.30	4.30

¹ the tabulated *t* value associated with an area of 0.025 in one tail of the *t*-distribution corresponding to the appropriate degrees of freedom, *v*, where *v* = *n*-1.

- 5 Table 4 summarizes the results of the statistical amino acid assignments for the 19 amino acids sequenced from the C-terminus of ACTH 7-38 fragment using the prior art time-dependent strategy. The masses of the listed amino acids could not be statistically differentiated from the experimentally derived mass difference at the given confidence levels. The amino acids indicated in bold are the known residues existing at the given positions. The confidence intervals indicated
- 10 are the highest levels at which all amino acid masses other than those indicated are statistically different from the experimental mean.

- 39 -

Table 4

ACTH 7-38 Fragment Amino Acid Position	Amino Acid Assignments ¹	Confidence Interval (c.i.)
20	Val	95% < c.i. < 98%
21	Gln/Lys	c.i. > 99.8%
22	Val	c.i. > 99.8%
23	Tyr	99% < c.i. < 99.8%
24	Pro	95% < c.i. < 98%
25	Asn	98% < c.i. < 99%
26	Gly	c.i. > 99.8%
27	Ala	98% < c.i. < 99%
28	Gln/Lys/Glu/Met	95% < c.i. < 98%
28	Met	80% < c.i. < 90%
29	Asp	99% < c.i. < 99.8%
30	Glu	c.i. > 99.8%
31	Ser	c.i. > 99.8%
32	Ala	c.i. > 99.8%
33	Glu	c.i. > 99.8%
34	Ala	c.i. > 99.8%
35	Phe	c.i. > 99.8%
36	Pro	99% < c.i. < 99.8%
37	Leu(Ile)/Asn	95% < c.i. < 98%
38	Gln/Lys/Glu	98% < c.i. < 99%
38	Gln/Lys	80% < c.i. < 90%

¹ assuming that only the 20 common unmodified amino acids are probable candidates

For example, the distinction between Gln and Lys for the amino acid assignment of residue 21 could not be made as the experimental mean (128.15 daltons) exactly bisected the asserted means of Gln (128.13 daltons) and Lys (128.17 daltons). The same phenomenon occurred in the assignment of residue 37. The experimental mean (113.63 daltons) bisected the asserted means of Leu(Ile) (113.16 daltons) and Asn (114.10 daltons). The assignments of the amino acids at positions 28 and 38 were difficult due to the small number of replicates taken (2 and 3, respectively). Residue 28 was assigned Gln/Lys/Glu/Met at a confidence interval greater than 95% but less than 98%. Table 3 shows that, for this residue, the asserted amino acid mass that resulted in the smallest $t_{calculated}$ was that of methionine. Using a confidence interval of 80%, the correct assignment of Glu is deemed statistically improbable. Likewise, the assignment of residue 38 was made as Gln/Lys/Glu at a confidence level of 95%, but the correct assignment (Glu) is again statistically improbable at an 80% level.

- 40 -

Since the errors are randomly distributed, all amino acids can be differentiated (except Leu and Ile) by sufficient population sampling. Approximating the experimental standard deviation to be that given above of $s = 0.604$ for the overall experiment, it is approximated (using $t_{table} = 1.960$) that >876 measurements would be required to differentiate Gln and Lys ($\Delta \text{mass} = 0.04$ daltons) at a 95% confidence interval. This number is experimentally impractical, but can be significantly lowered by reducing the standard deviation of the experimental mean. Decreasing the experimental standard deviation is of significant value as the number of samples required for the distinction between two amino acids to be made is proportional to the square of the experimental standard deviation of the mass difference. It is anticipated that mass shift reagents used to move peptide populations out of the interfering matrix are a possible chemical means for improving experimental error relating to peptides appearing in the low mass (<600 daltons) region. The use of reflectron and/or extended flight tube geometries are also expected to be instrumental methods suitable for reducing this error.

The protocol disclosed herein for statistical assignment of residues using the on-plate strategy involves multiple sampling from each well in which digestion is performed. The number of replicates required depends on the amino acid(s) that is(are) being sequenced at any one CPY concentration. For example, more replicates are required for mass differences around 113-115 daltons (Ile/Leu, Asn and Asp) and 128-129 daltons (Gln/Lys/Glu) than for mass differences around 163 (Tyr) or 57 (Gly) in order to be able to assure that all but one assignment are statistically unlikely. The experimental errors for this method appear to be as random (multiple replicates per sample) as for the time-dependent digestion (one replicate per sample).

This general statistical protocol for residue assignment was applied to two adjacent peaks that represent the loss of two or more amino acids. In this case, the asserted means of all dipeptides, tripeptides, etc. can also be used to calculate t-values. The information concerning the order of the residues will be lost but the composition can be deduced. Using only single amino acid and dipeptide masses as asserted means this was done for angiogenin has a sequence gap of Phe-Arg (Table 1). The average experimental mass difference between the peaks representing the loss of Arg(15) and Phe(13) was 303.45 ± 0.328 ($n=5$). For all single amino acid and dipeptide masses except Phe/Arg, the calculated t-values are greater than the tabulated t-value at a confidence interval of 99.8%. In this particular case, the identity of the amino acids that comprise the gap was determined, but their order remains experimentally unknown. This statistical strategy

- 41 -

was also incorporated into a computer algorithm to perform interactive data analysis and interpretation of ladder sequencing/MALDI experiments.

Thus, as illustrated above, the use of CPY digestion coupled with MALDI detection as disclosed herein was effective for obtaining C-terminal sequence information. The ACTH 7-38 fragment yielded sequence information 19 amino acids from the C-terminus without gaps. The on-plate concentration-dependent approach was demonstrated as a useful method for performing multiple digestions in parallel which circumvented the need for time- and reagent-consuming method development. This on-plate strategy required less physical manipulations and less total amounts of enzyme and peptide. Of the 22 peptides attempted using the on-plate approach, all but three were successfully digested to yield some C-terminal sequence information. CPY was also shown to cleave amidated C-terminal residues, but possessed no activity towards certain combinations of residues existing at the C-terminus and penultimate position.

In summary, an integrated strategy for generating residue assignments from "on-plate" C- and N-terminal peptide ladder sequencing experiments was developed. This strategy is based on the logical combination of tasks involving:

- 1) the creation of peptide ladders from a concentration-dependent exopeptidase digestion strategy that utilizes the μ L -wells of the Voyager™ sample plate as microreaction vessels;
- 2) the use of the Voyager™ MALDI-TOF workstation as a tool to generate masses of the peptide fragment;
- 3) an interpretation algorithm based on t-statistics that allows elimination of asserted assignment candidates; and,
- 4) feedback control of the data acquisition software from the interpretation algorithm that governs the number of replicates that are acquired for the statistically-based assignments to be made completely or to a cost effective partial point.

(c) Analysis of Experimentally-Obtained Mass-to-Charge Ratio Data: Nucleic Acids

The method disclosed herein has also been used to obtain sequence information about a nucleic acid polymer containing 40 bases. Hydrolysis using an exonuclease specific for the 3'

- 42 -

terminus was conducted using different concentrations of Phos I (phosphodiesterase I) ranging from 0.002 $\mu\text{U}/\mu\text{L}$ to 0.05 $\mu\text{U}/\mu\text{L}$. Hydrolysis was allowed to proceed for 3 minutes. Spectra of hydrolyzed sequences using MALDI-TOF are depicted in Figures 6A-6E. Data integration as disclosed herein confirmed the sequence to be:

5 CGC TCT CCC TTA TGC GAC TCC TGC ATT AGG AAG CAG CCC A (SEQ. ID. No. 23).

In a separate experiment, addition of a light-absorbent matrix CHCA was evaluated. A nucleic acid polymer containing 40 bases (as described above) was mixed with matrix and 0.4 $\mu\text{U}/\mu\text{L}$ of the exonuclease Phos II (phosphodiesterase II) which is specific for the 5' terminus. Hydrolysis in the presence of matrix was allowed to proceed for 10 minutes. The spectrum
10 obtained by MALDI-TOF is depicted in Figure 7. These data confirm the ability to combine polymer, hydrolyzing agent and matrix prior to mass spectrometry analysis. This reduces handling of reagents and facilitates sample processing. Using data similar to those in Figure 7, the sequence of the nucleic acid polymer was confirmed to be as described above.

EXAMPLE 4. OTHER APPLICATIONS OF THE INSTANT METHOD

15 As disclosed herein, this strategy can be applied to the sequencing of any natural biopolymer such as proteins, peptides, nucleic acids, carbohydrates, and modified versions thereof as well as synthetic biopolymers such as PNA and phosphothiolated nucleic acids. The ladders can be created enzymatically using exohydrolases, endohydrolases or the Sanger method and/or chemically by truncation synthesis or failure sequencing.

20 It is expected that other approaches can be taken to expand the utility of the CPY/MALDI ladder sequencing methods disclosed herein. For example, by taking advantage of different enzyme specificities, the use of carboxypeptidase mixtures can be implemented using the disclosed on-plate strategy as a means for sequencing through residue combinations that prohibit CPY activity as well as preventing sequence gaps from occurring. Also, by covalently attaching N-
25 terminal and/or C-terminal linkers to small peptides, it is expected that all sequence peaks can be made to fall beyond the low mass matrix region. It is anticipated that peptides can be completely sequenced to the N-terminus without gaps by combining MALDI with the above-described carboxypeptidase mixtures and mass shift reagent modifications.

- 43 -

Equivalents

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. The foregoing embodiments are therefore to be considered in a all respects illustrative rather than limiting on the invention described herein.

- 5 Scope of the invention is thus indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

- 44 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: PERSEPTIVE BIOSYSTEMS, INC.
- (B) STREET: 500 OLD CONNECTICUT PATH
- (C) CITY: FRAMINGHAM
- (D) STATE: MA
- (E) COUNTRY: USA
- (F) POSTAL CODE (ZIP): 01701
- (G) TELEPHONE: 508-383-7700
- (H) TELEFAX: 508-383-7852
- (I) TELEX:

(ii) TITLE OF INVENTION: METHODS AND APPARATUS FOR
SEQUENCING POLYMERS WITH A STATISTICAL CERTAINTY USING
MASS SPECTROMETRY

(iii) NUMBER OF SEQUENCES: 23

(iv) CORRESPONDENCE ADDRESS:

- (A) ADDRESSEE: PERSEPTIVE BIOSYSTEMS
- (B) STREET: 500 OLD CONNECTICUT PATH
- (C) CITY: FRAMINGHAM
- (D) STATE: MA
- (E) COUNTRY: USA
- (F) POSTAL CODE (ZIP): 01701

(v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/447,175
- (B) FILING DATE: 19-MAY-1995

- 45 -

(vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/446,055
- (B) FILING DATE: 19-MAY-1995

(viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: PITCHER, Edmund R.
- (B) REGISTRATION NUMBER: 27,829
- (C) REFERENCE/DOCKET NUMBER: SYP-122PC

(ix) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (617) 248-7000
- (B) TELEFAX: (617) 248-7100

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Trp Ala Gly Gly Asp Ala Ser Gly Glu
1 5

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Val His Leu Thr Pro Val Glu Lys
1 5

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 46 -

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Val Gln Gly Glu Glu Ser Asn Asp Lys

1

5

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 8 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Lys Arg Gln His Pro Gly Lys Arg

1

5

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Arg Pro Pro Gly Phe Ser Pro Phe Arg

1

5

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Glu His Trp Ser Tyr Gly Leu Arg Pro Gly

1

5

10

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- 47 -

- (A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
Glu Ala Asp Pro Asn Lys Phe Tyr Gly Leu Met
1 5 10
- (2) INFORMATION FOR SEQ ID NO:8:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
Asp Arg Val Tyr Ile His Pro Phe His Leu
1 5 10
- (2) INFORMATION FOR SEQ ID NO:9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
Pro His Pro Phe His Phe Phe Val Tyr Lys
1 5 10
- (2) INFORMATION FOR SEQ ID NO:10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
Asp Val Pro Lys Ser Asp Gln Phe Val Gly Leu Met
1 5 10

- 48 -

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Arg Pro Lys Pro Gln Gln Phe Phe Gly Leu Met
1 5 10

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Cys Gly Tyr Gly Pro Lys Lys Lys Arg Lys Val Gly Gly
1 5 10

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Gly Ala Pro Val Pro Tyr Pro Asp Pro Leu Glu Pro Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- 49 -

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Ala Asp Ser Gly Glu Gly Asp Phe Leu Ala Glu Gly Gly Gly Val Arg
1 5 10 15

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Gly Glu Gln Arg Lys Asp Val Tyr Val Gln Leu Tyr Leu
1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Glu Gln Arg Leu Gly Asn Gln Trp Ala Val Gly His Leu Met
1 5 10

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Lys Pro Val Gly Lys Lys Arg Arg Pro Val Lys Val Tyr Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids

- 50 -

- (B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
Ser Thr Ser Met Glu His Phe Arg Trp Gly Lys Pro Val
1 5 10
- (2) INFORMATION FOR SEQ ID NO:19:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
Asp Arg Val Tyr Ile His Pro Phe His Leu Leu Val Tyr Ser
1 5 10
- (2) INFORMATION FOR SEQ ID NO:20:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
Glu Asn Gly Leu Pro Val His Leu Asp Gln Ser Ile Phe Arg Arg
1 5 10 15
- (2) INFORMATION FOR SEQ ID NO:21:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: peptide
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
His Ser Gln Gly Thr Phe Thr Ser Asp Tyr Ser Lys Tyr Leu Asp Ser
1 5 10 15

- 51 -

Arg Arg Ala Gln Asp Phe Val Gln Trp Leu Met Asn Thr
 20 25

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

Phe Arg Trp Gly Lys Pro Val Gly Lys Lys Arg Arg Pro Val Lys Val
 1 5 10 15
 Tyr Pro Asn Gly Ala Glu Asp Glu Ser Ala Glu Ala Phe Pro Leu Glu
 20 25 30

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(A) DESCRIPTION: /desc = "NUCLEIC ACID POLYMER"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CGCTCTCCCT TATGCGACTC CTGCATTAGG AAGCAGCCCA

40

What is claimed is:

- 1 1. A method of obtaining sequence information about a polymer comprising a plurality of
2 monomers of known mass, said method comprising the steps of:
 - 3 a) providing a set of polymer fragments, each differing by one or more
4 monomers;
 - 5 b) measuring a difference x between the mass-to-charge ratio of at least one pair of
6 fragments;
 - 7 c) asserting a mean difference μ between the mass-to-charge ratio of the pair of
8 fragments measured in step b, wherein μ corresponds to a known mass-to-charge
9 ratio of one or more differing monomers;
 - 10 d) selecting a desired confidence level for μ ;
 - 11 e) analyzing x to determine if it is statistically different from μ by the selected
12 confidence level.
- 1 2. The method of claim 1 wherein a statistical difference determined in the analysis of step e)
2 indicates that the asserted mean μ is not assignable to the mass difference x with the
3 selected confidence level.
- 1 3. The method of claim 2 comprising repeating steps c) through e) until all desired μ s have
2 been asserted.
- 1 4. The method of claim 2 wherein the analysis of step e) comprises a two-tailed t-test for one
2 experimental mean.
- 1 5. The method of claim 1 wherein the analyzing in step e) comprises:
 - 2 f) repeating step b) a number of times, n , to determine a measured mean
3 mass-to-charge ratio difference \bar{x} between at least one pair of fragments;
 - 4 g) determining a standard deviation s of the mean mass-to-charge ratio
5 difference \bar{x} determined in step f);
 - 6 h) comparing \bar{x} to the asserted mean difference μ ;
 - 7 i) repeating steps c) through h) until all desired μ s have been asserted.

- 53 -

- 1 6. The method of claim 5 comprising repeating steps b) through i) for additional pairs of
2 fragments.
- 1 7. The method of claim 5 wherein the comparing in step h) is taking the absolute value of the
2 difference.
- 1 8. The method of claim 5 further comprising the step of determining the number of
2 measurements, n, based upon the analysis in step e).
- 1 9. The method of claim 1 wherein the polymer is a biopolymer.
- 1 10. The method of claim 9 wherein the biopolymer is selected from the group consisting of
2 DNAs, RNAs, PNAs, proteins, peptides, carbohydrates and modified forms thereof.
- 1 11. The method of claim 1 further comprising the step of hydrolyzing the polymer to obtain
2 the polymer fragments in step a).
- 1 12. The method of claim 1 further comprising hydrolyzing, on a reaction surface, the polymer
2 with a hydrolyzing agent.
- 1 13. The method of claim 12 wherein the polymer is hydrolyzed on a reaction surface, said
2 surface providing differing amounts of a hydrolyzing agent which hydrolyzes said polymer
3 thereby to break inter-monomer bonds.
- 1 14. The method of claim 11, 12 or 13 wherein the hydrolyzing agent is an exohydrolase or an
2 endohydrolase.
- 1 15. The method of claim 14 wherein hydrolyzing with said exohydrolase produces a series of
2 fragments comprising a sequence-defining ladder of said polymer.
- 1 16. The method of claim 15 wherein the exohydrolase is selected from the group consisting
2 of: exonucleases, exoglycosidases, and exopeptidases.
- 1 17. The method of claim 16 wherein the exopeptidase is selected from the group consisting of
2 carboxypeptidase Y, carboxypeptidase A, carboxypeptidase B, carboxypeptidase P,
3 aminopeptidase 1, leucine aminopeptidase, proline aminodipeptidase and cathepsin C.

- 54 -

- 1 18. The method of claim 16 wherein the exoglycosidase is selected from the group consisting
2 of
3 a) α - Mannosidase I
4 b) α - Mannosidase
5 c) β - Hexosaminidase
6 d) β - Galactosidase
7 e) α - Fucosidase I and II
8 f) α - Galactosidase
9 g) α - Neuraminidase
10 h) α - Glucosidase I and II.
- 1 19. The method of claim 16 wherein the exonuclease is selected from the group consisting of
2 a) Exonuclease
3 b) λ - exonuclease
4 c) t7 Gene 1 exonuclease
5 d) exonuclease III
6 e) Exonuclease I
7 f) Exonuclease V
8 g) Exonuclease II
9 h) DNA Polymerase II
- 1 20. The method of claim 14 wherein hydrolyzing with said endohydrolase produces a series of
2 fragments defining a map of said polymer.
- 1 21. The method of claim 20 wherein said endohydrolase is an endopeptidase selected from the
2 group consisting of: trypsin, chymotrypsin, endo-proteinase Lys-C, endoproteinase Arg-C
3 and thermolysin.
- 1 22. The method of claim 12 wherein the agent is a hydrolyzing agent other than an enzyme.
- 1 23. The method of claim 12 wherein said agent capable of hydrolyzing said polymer comprises
2 a combination of at least one enzyme and at least one agent other than an enzyme.
- 1 24. The method of claim 13 wherein the reaction surface comprises an array of discrete
2 separable zones, each zone comprising a differing amount of said hydrolyzing agent.

- 55 -

- 1 25. The method of claim 13 wherein the reaction surface comprises a non-discrete gradient of
2 said hydrolyzing agent.
- 3 26. The method of claim 12 wherein said reaction surface comprises a constant amount of said
4 polymer.
- 1 27. The method of claim 12 wherein said reaction surface comprises an array of discrete
2 separate zones of differing amounts of said polymer.
- 1 28. The method of claim 12 wherein said reaction surface comprises a non-discrete gradient of
2 said polymer.
- 1 29. The method of claim 12 wherein said reaction surface comprises a constant amount of said
2 agent.
- 1 30. The method of claim 1 further comprising adding a matrix to the polymer fragments
2 before measuring the mass-to-charge ratio in step b).
- 1 31. The method of claim 1 wherein the ratio is analyzed by matrix assisted laser desorption
2 mass spectrometry.
- 1 32. The method of claim 1 wherein step (b) is conducted by plasma desorption ionization or
2 fast atom bombardment ionization.
- 1 33. The method of claim 1 wherein step (b) is accomplished using mass analysis modes
2 selected from the group consisting of: time-of-flight, quadrapole, ion trap, and sector.
- 1 34. The method of claim 12 wherein said reaction surface comprises a mass spectrometer
2 sample holder having microreaction vessels disposed thereon.
- 1 35. The method of claim 12 wherein said reaction surface comprises a mass spectrometer
2 sample probe.
- 1 36. The method of claim 12 wherein said reaction surface comprises a substrate selected from
2 the group consisting of: metals, foils, plastics, ceramics, and waxes.
- 1 37. The method of claim 12 wherein hydrolysis is accomplished with dehydrated hydrolyzing
2 agent on said reaction surface.

- 56 -

38. The method of claim 12 wherein hydrolysis is accomplished by immobilizing said agent on said reaction surface.
39. The method of claim 12 wherein hydrolysis is accomplished using a hydrolyzing agent in liquid or gel form, said liquid or gel form being resistant to physical dislocation.
40. The method of claim 1 comprising the additional step of combining a light-absorbent matrix with said fragments prior to step b).
41. The method of claim 1 comprising the additional step of combining said polymer fragments with moieties for selectively shifting the mass of hydrolyzed sequences prior to step b).
42. The method of claim 1 comprising the additional step of combining said polymer fragments with moieties for improving ionization prior to step b).
43. A method for obtaining sequence information about a polymer comprising a series of different monomers of known mass, said method comprising the steps of:
- a) providing a set of polymer fragments, each differing by one or more monomers;
 - b) measuring the mass-to-charge ratio difference x between a pair of fragments;
 - c) asserting a mean difference μ , which is related to a known mass-to-charge ratio of one or more monomers;
 - d) selecting a desired confidence level for μ ;
 - e) repeating step b) to obtain a number of measurements n , thereby to determine the measured mean mass-to-charge ratio difference \bar{x} between the pair of fragments;
 - f) determining the standard deviation s of the measured mean mass-to-charge ratio difference \bar{x} determined in step e);
 - g) calculating a test statistic $t_{\text{calculated}}$ with the following algorithm:

$$t_{\text{calculated}} = \frac{|\bar{x} - \mu| \sqrt{n}}{s}$$

- 57 -

- 1 44. The method of claim 43 further comprising a comparison of the calculated test statistic
2 $t_{\text{calculated}}$ in step g) to a t-distribution corresponding to the number of measurements and the
3 desired confidence level.
- 1 45. The method of claim 43 further comprising repeating steps b)- g) for additional pairs of
2 fragments thereby to obtain sequence information.
- 1 46. The method of claim 44 further comprising the step of determining the number of
2 measurements, n , based upon the comparison.
- 1 47. The method of claim 43 wherein the polymer is a biopolymer.
- 1 48. The method of claim 47 wherein the biopolymer is selected from the group consisting of
2 DNAs, RNAs, PNAs, proteins, peptides, carbohydrates and modified forms thereof.
- 1 49. The method of claim 43 further comprising the step of hydrolyzing the polymer with a
2 hydrolyzing agent to create the fragments in step a).
- 1 50. The method of claim 49 wherein the hydrolyzing agent is an exohydrolase which
2 produces a series of fragments comprising a sequence-defining ladder of said polymer.
- 1 51. The method of claim 50 wherein the exohydrolase is selected from the group consisting
2 of: exonucleases, exoglycosidases, exopeptidases,
- 1 52. The method of claim 51 wherein the exopeptidase is selected from the group consisting of
2 carboxypeptidase Y, carboxypeptidase A, carboxypeptidase B, carboxypeptidase P,
3 aminopeptidase 1, leucine aminopeptidase, proline aminodipeptidase and cathepsin C.
- 1 53. The method of claim 51 wherein the exoglycosidase is selected from the group consisting
2 of
- 3 a) α - Mannosidase I
4 b) α - Mannosidase
5 c) β - Hexosaminidase
6 d) β - Galactosidase
7 e) α - Fucosidase I and II
8 f) α - Galactosidase

- 58 -

- 9 g) α - Neuraminidase
10 h) α - Glucosidase I and II.
- 1 54. The method of claim 51 wherein the exonuclease is selected from the group consisting of
2 a) Exonuclease
3 b) λ - exonuclease
4 c) t7 Gene 1 exonuclease
5 d) exonuclease III
6 e) Exonuclease I
7 f) Exonuclease V
8 g) Exonuclease II
9 h) DNA Polymerase II
- 1 55. The method of claim 49 wherein the hydrolyzing agent is other than an enzyme.
- 1 56. The method of claim 49 wherein the agent comprises a combination of at least one enzyme
2 and at least one agent other than an enzyme.
- 1 57. The method of claim 49 wherein hydrolysis is performed on a reaction surface, said
2 surface providing differing amounts of a hydrolyzing agent.
- 1 58. The method of claim 57 wherein the reaction surface comprises an array of discrete
2 separable zones, each zone comprising a differing amount of said hydrolyzing agent.
- 1 59. The method of claim 49 wherein the reaction surface comprises a continuous
2 concentration gradient of a hydrolyzing agent.
- 1 60. The method of claim 43 further comprising adding a matrix to the polymer fragments
2 before measuring the mass-to-charge ratio in step b).
- 1 61. A method for obtaining sequence information about a polymer having a plurality of
2 monomers of known mass, said method comprising:
3 a) providing a set of polymer fragments, each differing by one or more monomers;
4 b) measuring a difference x between the mass-to-charge ratio of a pair of fragments;
5 c) asserting a mean difference μ , which is related to the mass-to-charge ratio of the
6 pair of fragments measured in step b), wherein μ corresponds to the known mass-

- 59 -

- 7 to-charge ratio of one or more monomers;
- 8 d) selecting the desired confidence level for μ ;
- 9 e) analyzing x to determine if it is statistically different from μ by the selected
- 10 confidence level;
- 11 f) repeating steps b)-e) a number of times n , until all desired μ s have been asserted;
- 12 g) repeating steps b) -f) for additional pairs of fragments.

1 62. The method of claim 61 wherein the polymer is a biopolymer.

1 63. The method of claim 62 wherein the biopolymer is selected from the group consisting of

2 DNAs, RNAs, PNAs, proteins, peptides, carbohydrates and modified forms thereof.

1 64. The method of claim 61 wherein the polymer fragments in step a) are created by

2 concentration dependent hydrolysis of the polymer.

1 65. The method of claim 61 further comprising the step of hydrolyzing said polymer with a

2 hydrolyzing agent to produce the polymer fragments in step a).

1 66. The method of claim 65 wherein the hydrolyzing agent is an exohydrolase.

1 67. The method of claim 66 wherein the hydrolysis caused by said exohydrolase produces a

2 series of fragments defining a ladder of said polymer.

1 68. The method of claim 66 wherein the exohydrolase is selected from the group consisting

2 of: exonucleases, exoglycosidases, and exopeptidases.

1 69. The method of claim 68 wherein the exoglycosidase is selected from the group consisting

2 of

- 3 a) α - Mannosidase I
- 4 b) α - Mannosidase
- 5 c) β - Hexosaminidase
- 6 d) β - Galactosidase
- 7 e) α - Fucosidase I and II
- 8 f) α - Galactosidase
- 9 g) α - Neuraminidase
- 10 h) α - Glucosidase I and II.

- 60 -

- 1 70. The method of claim 68 wherein the exonuclease is selected from the group consisting of
2 a) Exonuclease
3 b) λ - exonuclease
4 c) t7 Gene 1 exonuclease
5 d) exonuclease III
6 e) Exonuclease I
7 f) Exonuclease V
8 g) Exonuclease II
9 h) DNA Polymerase II
- 1 71. The method of claim 68 wherein the exopeptidase is selected from the group consisting of
2 carboxypeptidase Y, carboxypeptidase A, carboxypeptidase B, carboxypeptidase P,
3 aminopeptidase 1, leucine aminopeptidase, proline, aminodipeptidase and cathepsin C.
- 1 72. The method of claim 65 wherein said agent comprises a hydrolyzing agent other than an
2 enzyme.
- 1 73. The method of claim 65 wherein the polymer fragments are obtained by hydrolysis with a
2 combination of at least one enzyme and at least one hydrolyzing agent other than an
3 enzyme.
- 1 74. The method of claim 65 wherein the hydrolysis occurs on a reaction surface, said surface
2 providing differing amounts of a hydrolyzing agent.
- 1 75. The method of claim 74 wherein the reaction surface comprises an array of discrete
2 separable zones, each zone comprising a differing amount of a hydrolyzing agent.
- 1 76. The method of claim 74 wherein the reaction surface comprises a concentration gradient
2 of said hydrolyzing agent.
- 1 77. The method of claim 61 further comprising adding a matrix to the polymer fragments
2 before measuring the mass-to-charge ratio in step b).
- 1 78. An apparatus for obtaining sequence information about a polymer having a plurality of
2 monomers of known mass, said apparatus comprising:
3 a) a mass spectrometer having a sample plate which holds a set of polymer fragments,

- 61 -

- 4 each differing by one or more monomers; and
- 5 b) a computer responsive to the mass spectrometer for:
- 6 i) determining the mass-to-charge ratio difference x between a pair of
- 7 polymer fragments;
- 8 ii) asserting a mean difference μ between the mass-to-charge ratio of the pair
- 9 of fragments determined in step i), wherein μ corresponds to the known
- 10 mass-to-charge ratio of one or more monomers;
- 11 iii) analyzing x to determine if it is statistically different from μ with a desired
- 12 confidence level, wherein a statistical difference indicates that the asserted
- 13 mean μ is not assignable to x with the desired confidence level; and
- 14 iv) repeating steps ii) - iii) until all desired μ s have been asserted; and
- 15 v) repeating steps i) - iv) on additional pairs of fragments.

1 79. The apparatus of claim 78 wherein the computer determines the asserted mass-to-charge

2 ratio difference between pairs of polymer fragments.

1 80. The apparatus of claim 78 wherein the sample plate comprises a reaction surface which

2 provides differing amounts of a hydrolyzing agent which hydrolyzes said polymer thereby

3 to break inter-monomer bonds.

1 81. The apparatus of claim 80 wherein said reaction surface comprises an array of discrete

2 separate zones of differing amounts of said agent or a non-discrete gradient of said agent.

1 82. The apparatus of claim 80 wherein said reaction surface comprises a gradient of said

2 agent.

1 83. The apparatus of claim 78 further comprising a light-absorbent matrix suitable for matrix-

2 assisted laser desorption mass spectrometry.

1 84. An apparatus for obtaining sequence information about a polymer having a plurality of

2 monomers of known mass comprising:

- 3 A. a mass spectrometer comprising:
- 4 a) means for generating ions;
- 5 b) means for accelerating ions;
- 6 c) means for detecting ions; and

- 62 -

B. a computer responsive to the mass spectrometer comprising:

- d) means for determining the mass-to-charge ratio difference x between a pair of polymer fragments;
- e) means for asserting a mean difference μ between the mass-to charge ratio of the pair of fragments, wherein μ corresponds to a known mass-to-charge ratio of one or more monomers;
- f) means for analyzing x to determine if it is statistically different from μ with a desired confidence level;
- g) and means for determining when the desired number of possible μ s has been asserted.

85. A kit for obtaining sequence information by mass spectrometry about a polymer comprising one or more monomers of known mass, wherein said kit comprises:

- a) a mass spectrometry sample plate which holds a set of polymer fragments, each differing by one or more monomers; and
- b) a computer readable disc for rendering a computer responsive to the mass spectrometer for:
 - i) determining the mass-to-charge ratio difference x between at least one pair of polymer fragments;
 - ii) analyzing the mass-to-charge ratio differences of pairs of polymer fragments determined in step i) to determine if they statistically differ with a desired confidence level from an asserted mass-to-charge ratio difference μ , wherein μ corresponds to a known mass-to-charge ratio difference, and, wherein a statistical difference indicates that the μ is not assignable to x ;
 - iii) repeating steps i) to ii).

86. The kit of claim 85 wherein the sample plate comprises a reaction surface, said surface providing differing amounts of a hydrolyzing agent to hydrolyze said polymer into said fragments.

87. The kit of claim 85 wherein the sample plate further comprises a matrix suitable for matrix-assisted laser desorption mass spectrometry.

- 63 -

- 1 88. A computer readable disc for rendering a computer responsive to a mass spectrometer for:
- 2 i) determining the mass-to-charge ratio difference x between at least one pair of
- 3 polymer fragments generated from a polymer having a plurality of monomers, each
- 4 fragment differing by one or more monomers;
- 5 ii) analyzing the mass-to-charge ratio difference to determine if x statistically differs
- 6 from an asserted mass-to-charge ratio difference by a predetermined confidence
- 7 interval, and
- 8 iii) repeating step ii) for additional asserted mass-to-charge ratios;
- 9 iv) repeating steps i) to ii) for additional pairs of fragments.
- 1 89. A computer responsive to a mass spectrometer comprising:
- 2 a) means for determining the mass-to-charge ratio difference x between at least one
- 3 pair of sequence-defining polymer fragments generated from a polymer having a
- 4 plurality of monomers, each fragment differing by one or more monomers;
- 5 b) means for analyzing the mass-to-charge ratio difference to determine x statistically
- 6 differs from an asserted mass-to-charge ratio difference by a predetermined
- 7 confidence interval, and
- 8 c) means for repeating step b) until all desired asserted differences have been
- 9 asserted; and
- 10 d) means for repeating steps a) - c) until sequence information is obtained.
- 1 90. A computer responsive to a mass spectrometer comprising:
- 2 a) means for determining the mass-to-charge ratio difference x between at least one
- 3 pair of sequence-defining polymer fragments generated from a polymer having a
- 4 plurality of monomers, each fragment differing by one or more monomers;
- 5 b) means for analyzing the mass-to-charge ratio difference to determine x statistically
- 6 differs from an asserted mass-to-charge ratio difference by a predetermined
- 7 confidence interval, and
- 8 c) means for repeating step b) until all desired asserted differences have been
- 9 asserted; and
- 10 d) means for repeating steps a) - c) until sequence information is obtained.

- 64 -

- 1 91. The method of claim 90 wherein steps b) - f) are repeated for additional fragments until
2 information is obtained about the identity of the polymer with the desired confidence level
3 until sequence information is obtained.
- 1 92. The method of claim 90 wherein the hypothetical identity in step c) corresponds to a
2 known identity derived from a computer database of known sequences.
- 1 93. The method of any one of claims 1, 43, 61 or 90 further comprising the step of eluting
2 from a liquid chromatography column a sample comprising polymer fragments for which
3 sequence information is to be obtained.
- 1 94. The method of claim 93 wherein the sample eluted from the column is rendered
2 compatible with a mass spectrometer by contact with a buffer prior to step b).
- 1 95. The method of claims 1, 43, 61 or 90 wherein step a) further comprises the steps of:
2 (1) on a reaction surface, providing at least
3 (i) one amount of hydrolyzing agent which hydrolyzes said polymer thereby to
4 break intermonomer bonds and produce said set of polymer fragments, and
5 (ii) a sample of said polymer to form differing ratios of agent to polymer on
6 said reaction surface;
7 (2) incubating the product of step (1) for a time sufficient to obtain a plurality of series
8 of hydrolyzed polymer fragments; and
9 (3) performing mass spectrometry on a plurality of said series to obtain mass-to-
10 change ratio data for hydrolyzed polymer fragments contained herein.
- 1 96. The apparatus of claim 78 wherein said sample plate comprises a planar solid surface
2 having disposed therein at least one amount of a dehydrated agent capable of hydrolyzing
3 a polymer.
- 1 97. The apparatus of claim 78 wherein said sample plate comprises a planar solid surface
2 having disposed thereon at least one amount of an immobilized agent capable of
3 hydrolyzing a polymer.

- 65 -

- 1 98. The apparatus of claim 78 wherein said sample plate comprises a planar solid surface
2 having disposed thereon at least one amount of a hydrolyzing agent in liquid or gel form,
3 said liquid or gel form being resistant to physical dislocation.
- 1 99. For use with a mass spectrometry apparatus to adapt said apparatus for obtaining
2 sequence information about a polymer comprising a series of different monomers, a mass
3 spectrometer sample plate comprising a planar solid surface having disposed thereon at
4 least one amount of a dehydrated agent capable of hydrolyzing a polymer.
- 1 100. For use with a mass spectrometry apparatus to adapt said apparatus for obtaining
2 sequence information about a polymer comprising a series of different monomers, a mass
3 spectrometer sample plate comprising a planar solid surface having disposed thereon at
4 least one amount of an immobilized agent capable of hydrolyzing a polymer.
- 1 101. For use with a mass spectrometry apparatus to adapt said apparatus for obtaining
2 sequence information about a polymer comprising a series of different monomers, a mass
3 spectrometer sample plate comprising a planar solid surface having disposed thereon at
4 least one amount of a hydrolyzing agent in liquid or gel form, said liquid or gel form being
5 resistant to physical dislocation.
- 1 102. The sample plate of any one of claims 78, 85, 99, 100 or 101 wherein said surface
2 comprises an array of discrete separate zones of differing amounts of said agent.
- 1 103. The sample plate of any one of claims 78, 85, 99, 100 or 101 wherein said surface
2 comprises a non-discrete gradient of said agent.
- 1 104. The sample plate of any one of claims 78, 85, 99, 100 or 101 wherein said surface
2 comprises a constant amount of said agent.
- 1 105. The sample plate of any one of claims 78, 85, 99, 100 or 101 further comprising a light-
2 absorbent matrix.
- 1 106. The sample plate of any one of claims 78, 85, 99, 100 or 101 further comprising
2 microreaction vessels.

- 66 -

- 1 107. ... The sample plate of any one of claims 78, 85, 99, 100 or 101 wherein said plate is
- 2 disposable.

1/16

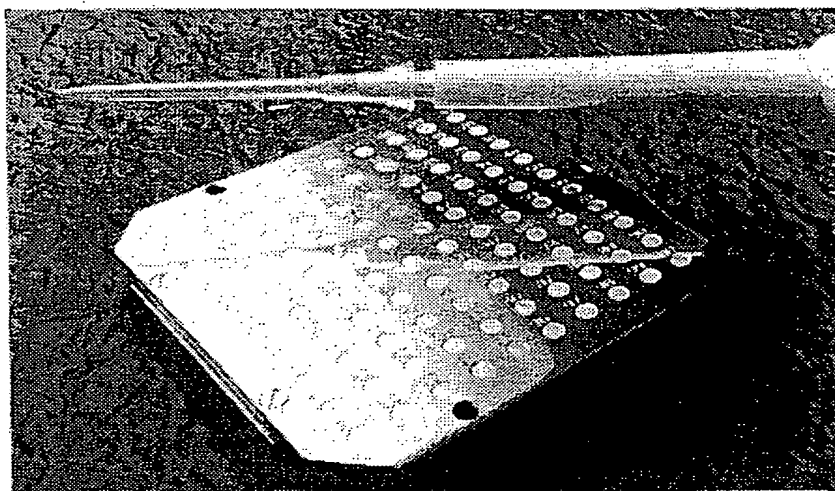
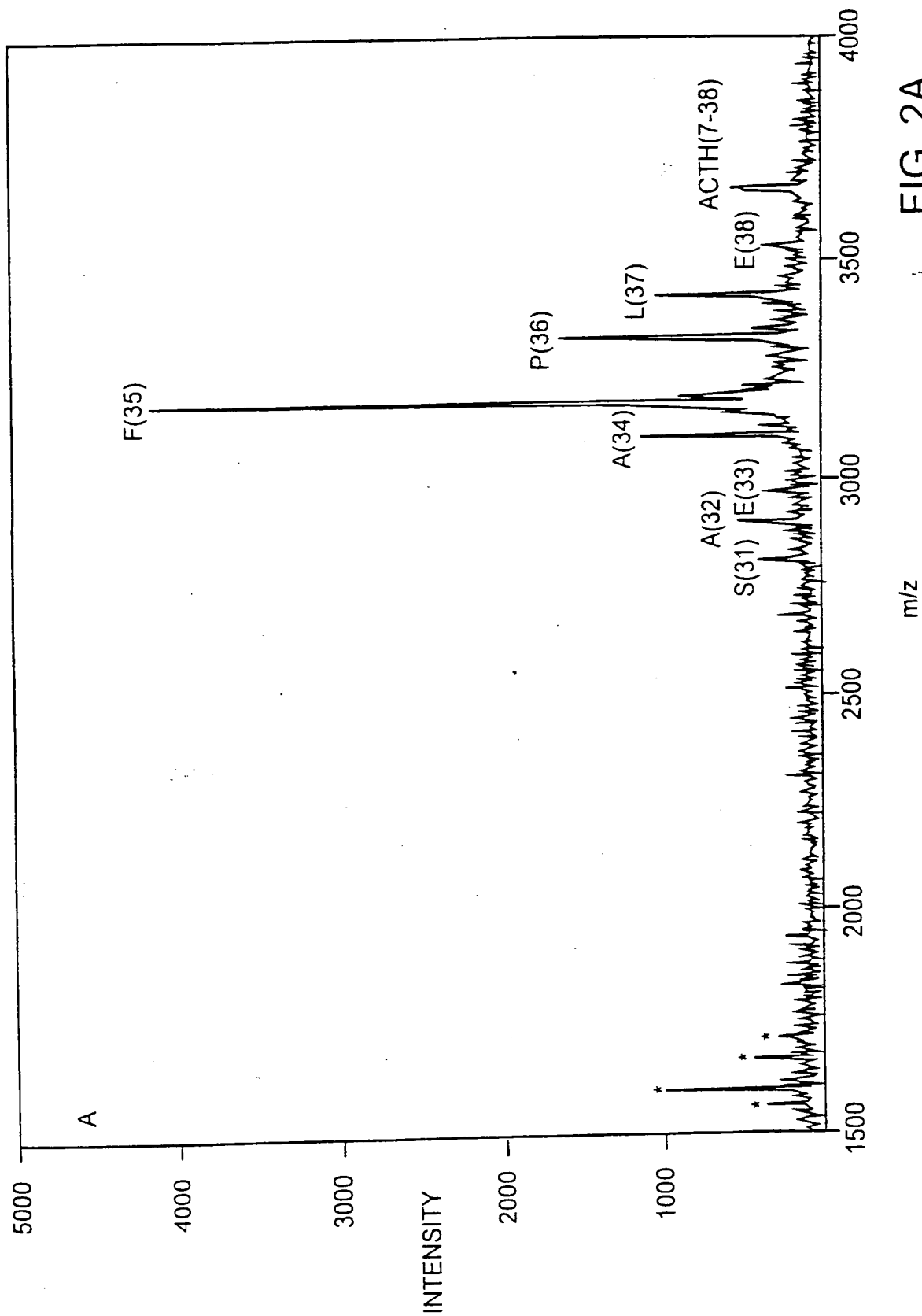


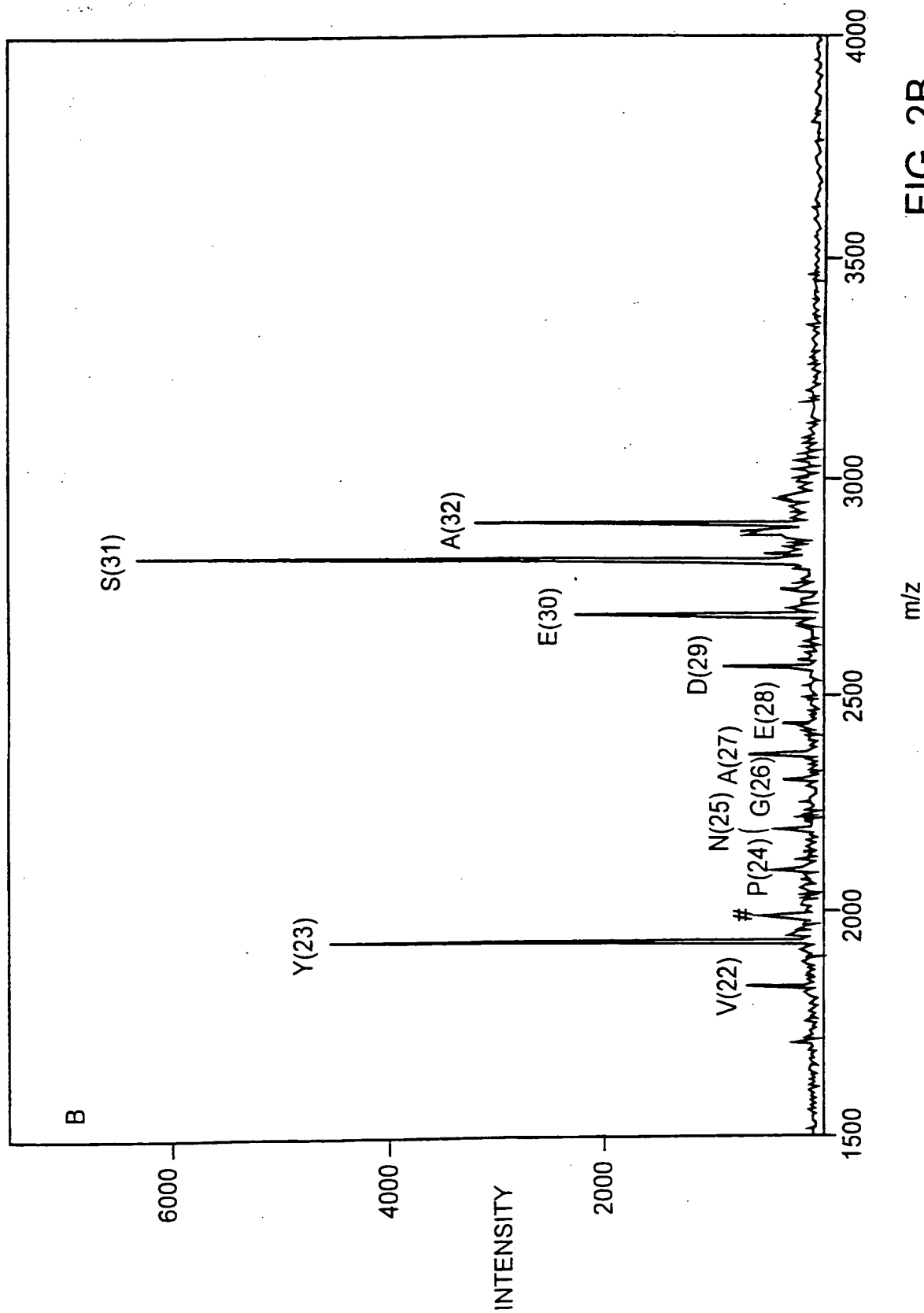
FIG. 1

2/16

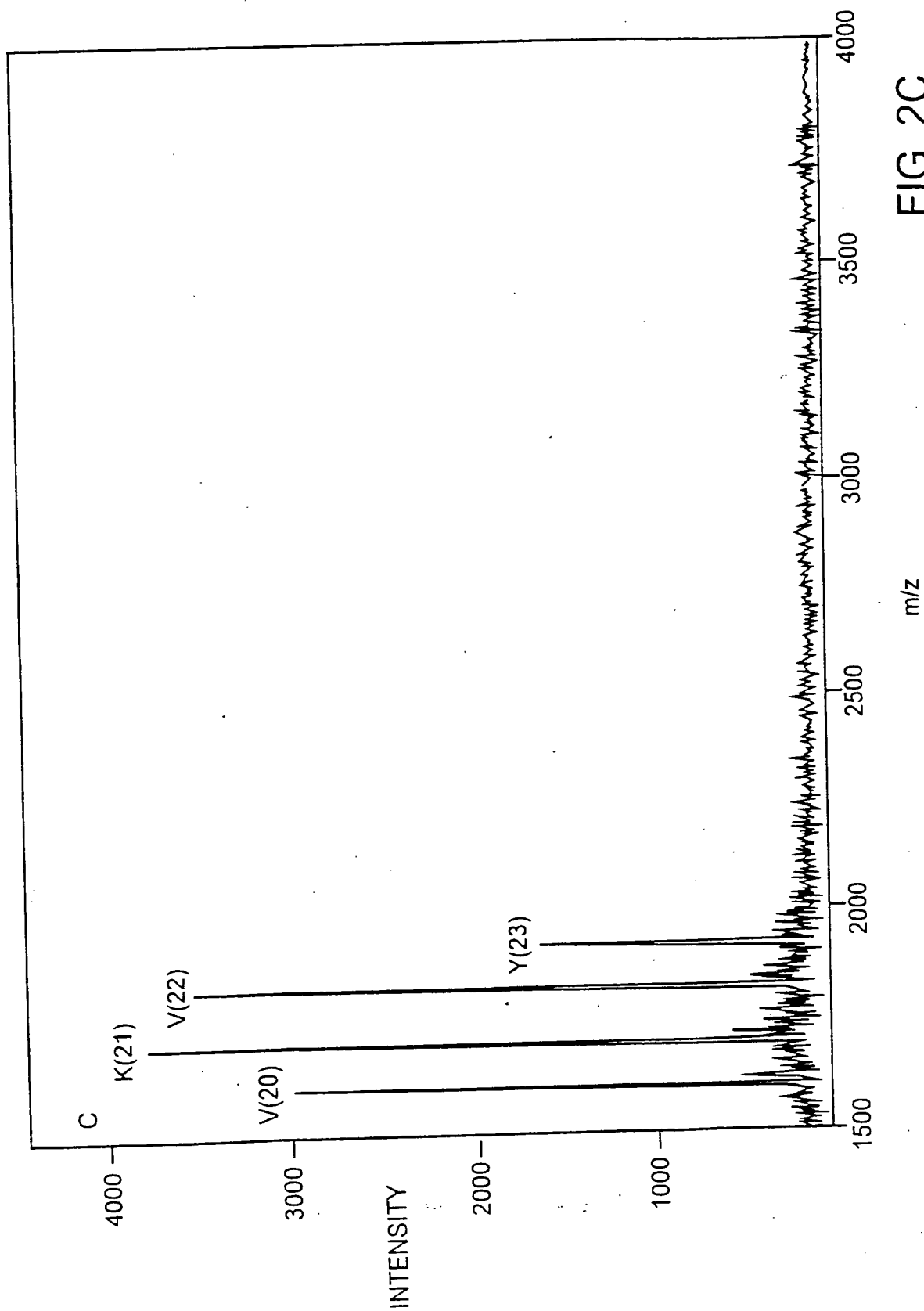


SUBSTITUTE SHEET (RULE 26)

3/16



4/16



SUBSTITUTE SHEET (RULE 26)

5/16

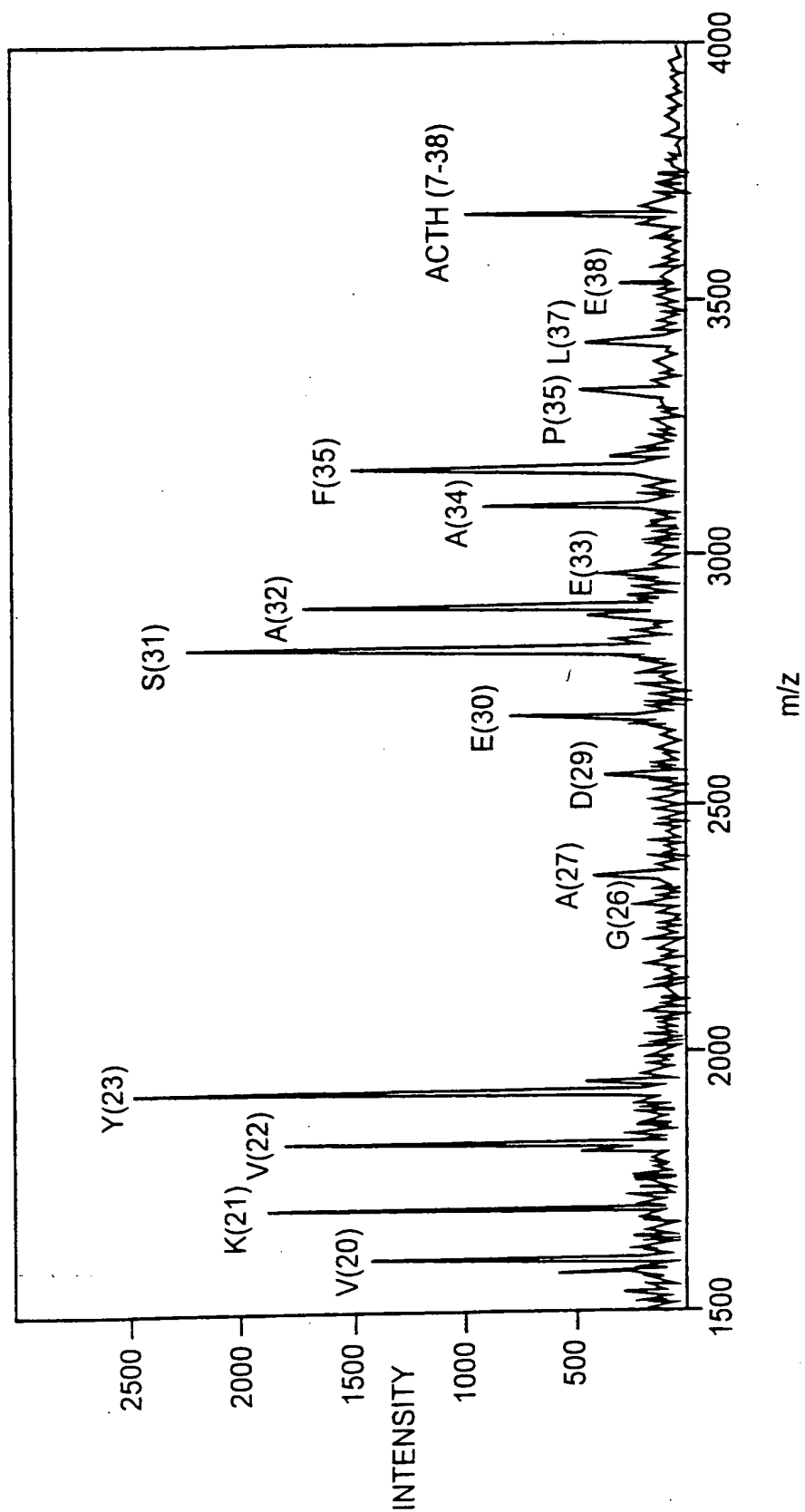
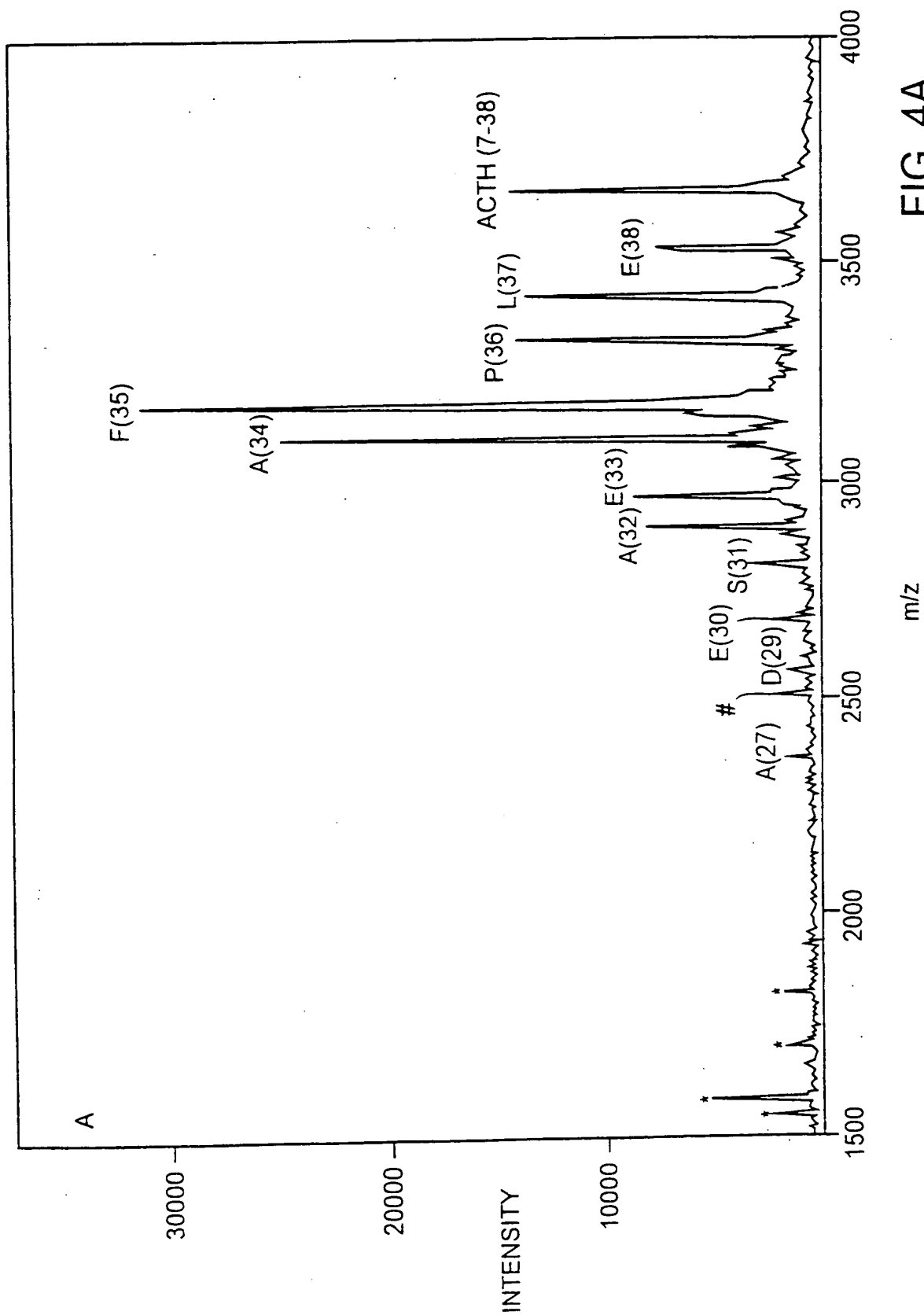


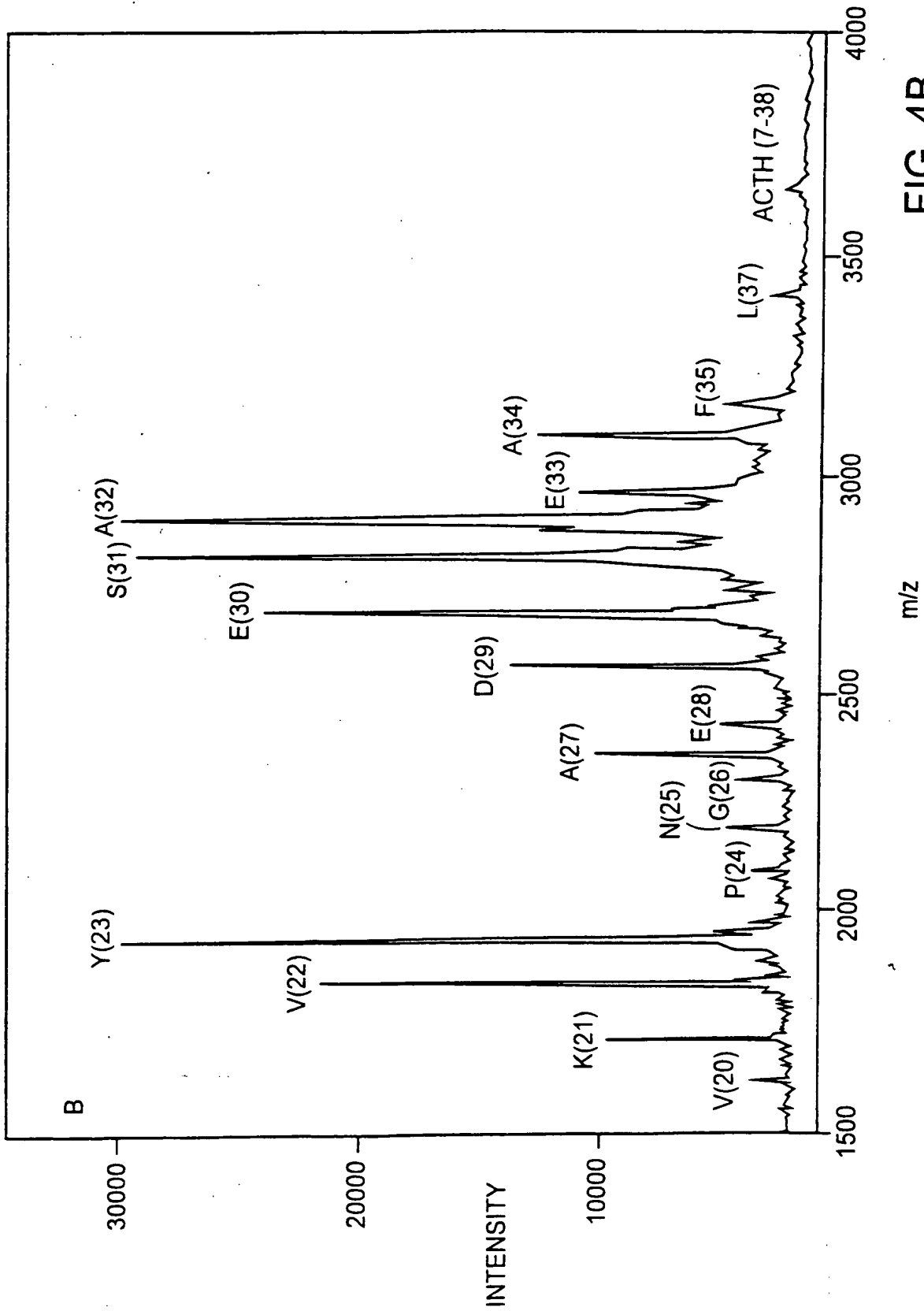
FIG. 3

6/16



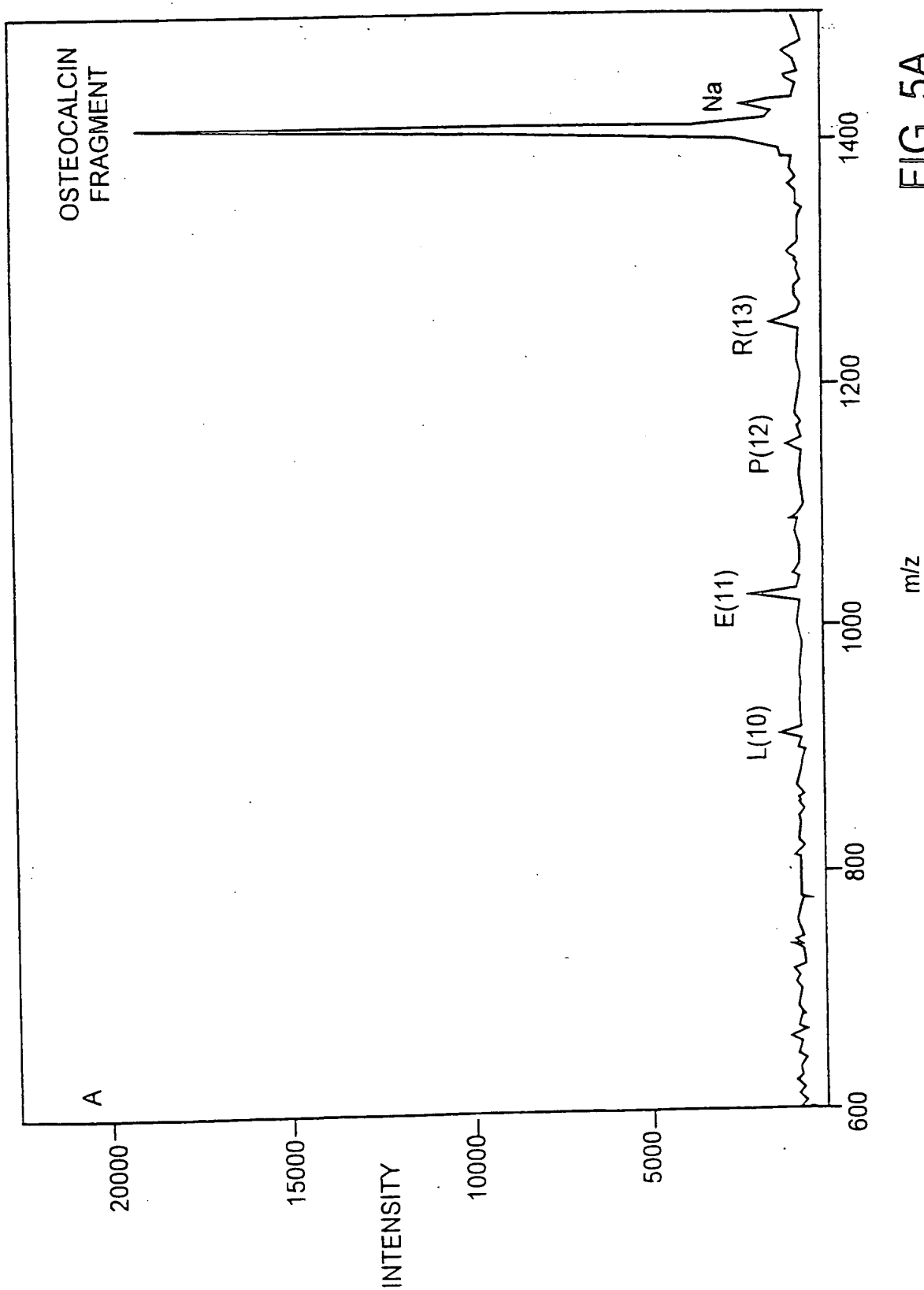
SUBSTITUTE SHEET (RULE 26)

7/16



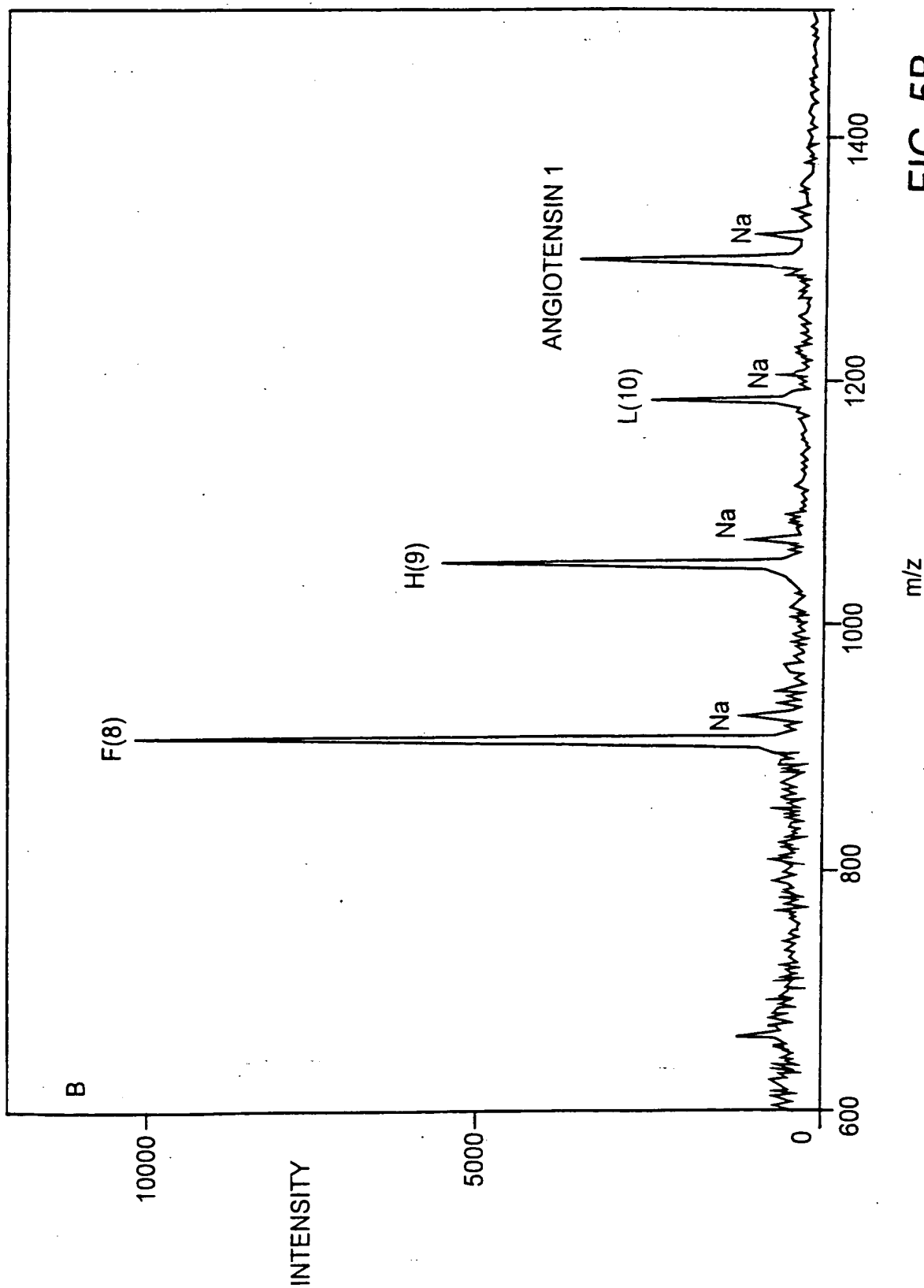
SUBSTITUTE SHEET (RULE 26)

8/16



SUBSTITUTE SHEET (RULE 26)

9/16



SUBSTITUTE SHEET (RULE 26)

10/16

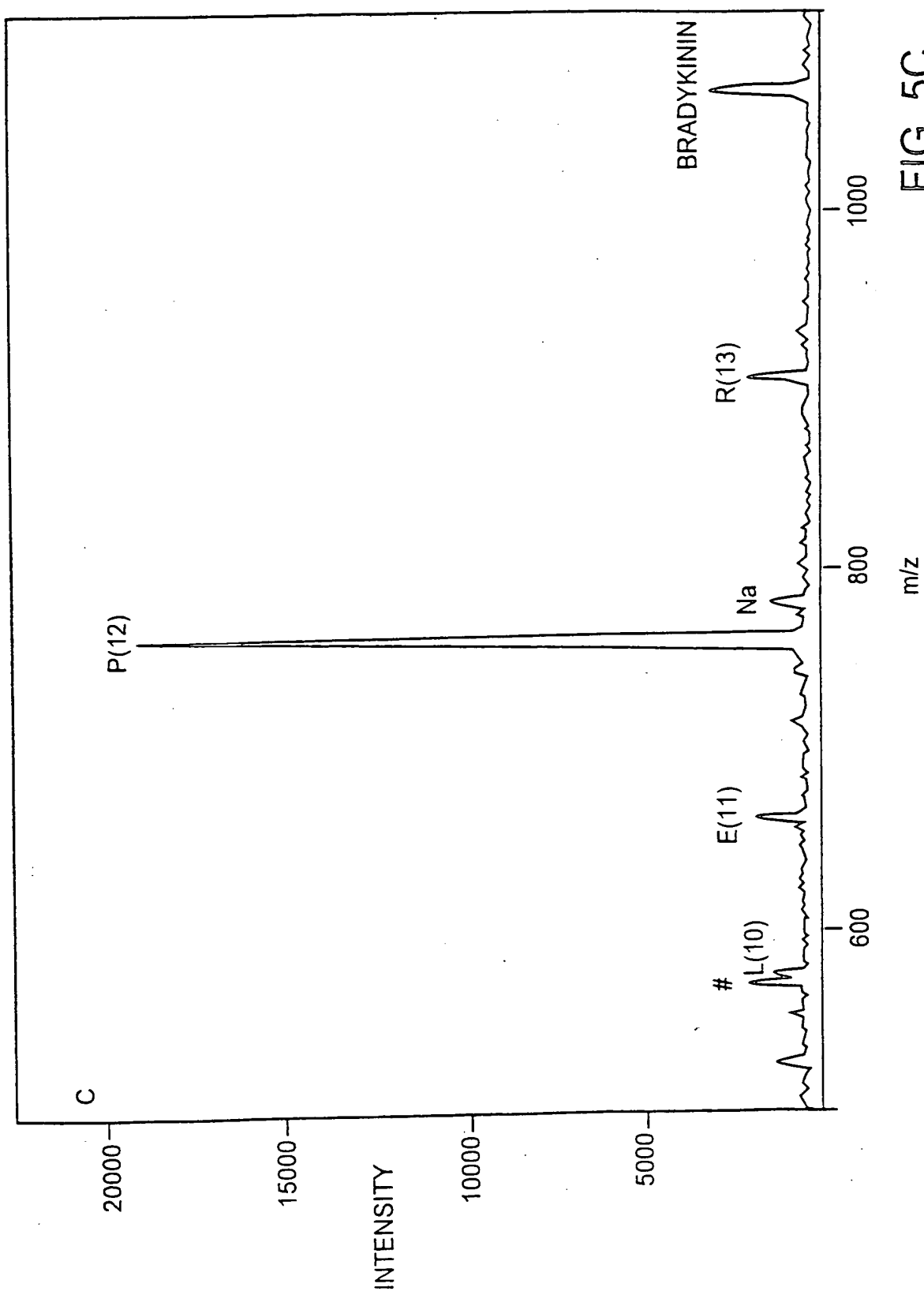
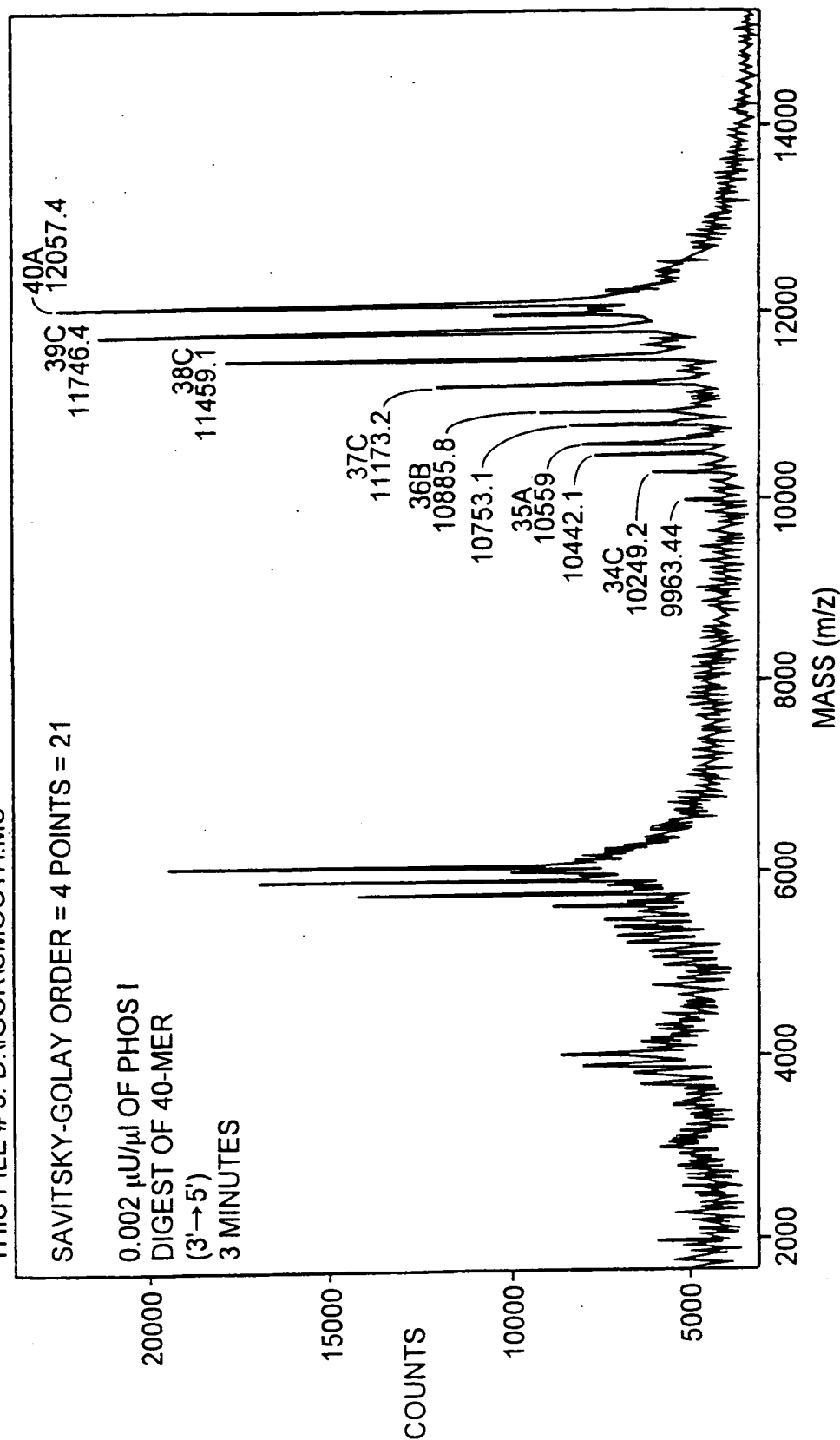


FIG. 5C

SUBSTITUTE SHEET (RULE 26)

11/16

ORIGINAL FILENAME: D:\IGOR\5427007.MS
THIS FILE # 3: D:\IGOR\SMOOTH.MS

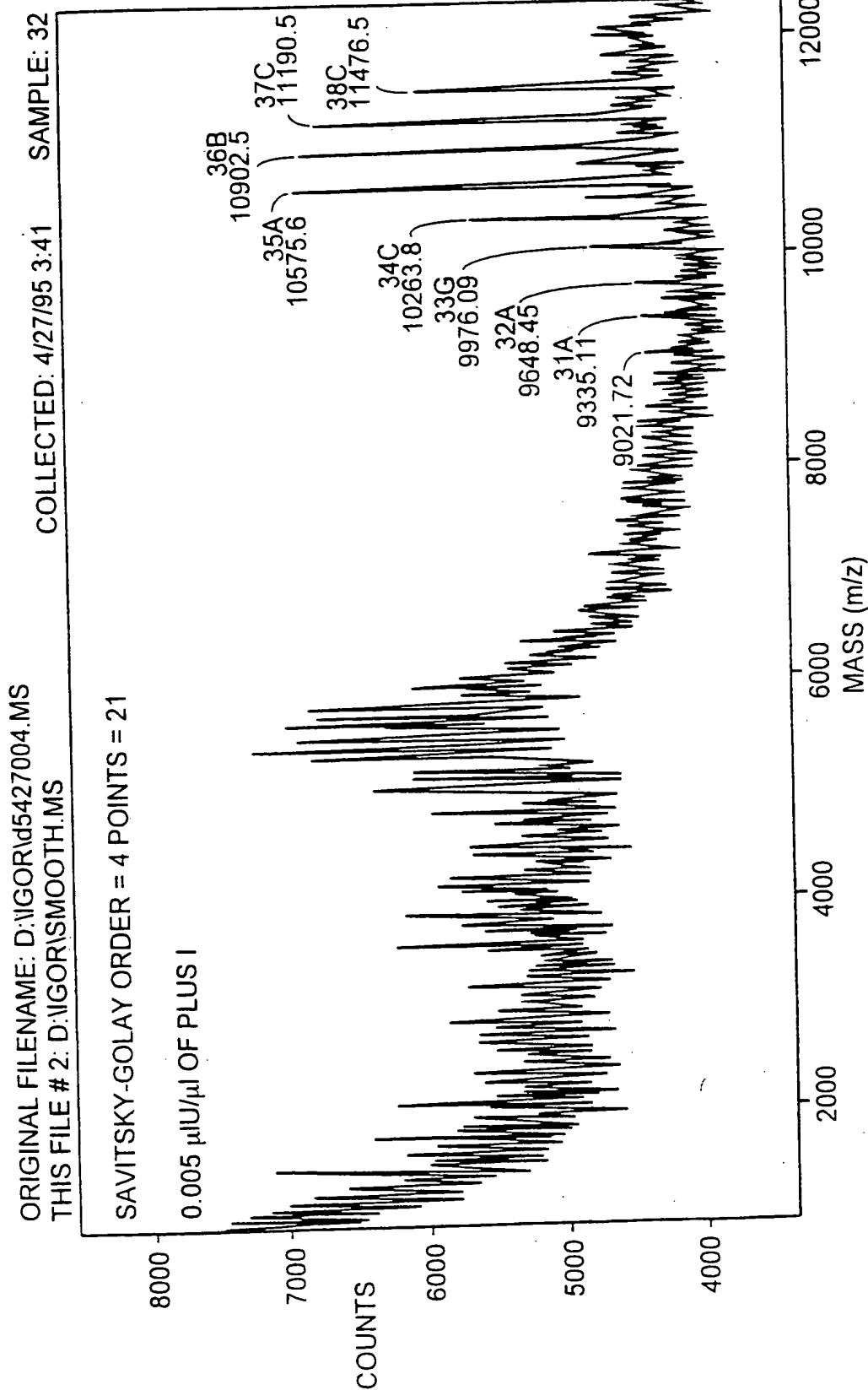


COMMENT: 51302, PHOS I, 3', 40-MER(h-1)

METHOD: DNA1	ACCELERATING VOLTAGE: 20000	LASER: 0	LOW MASS GATE: 800.0
MODE: LINEAR	GRID VOLTAGE: 93.0%	SCANS AVERAGED: 23	TIMED ION SELECTOR:
GUIDE WIRE VOLTAGE: 0.200%	PRESSURE: 3.90e-70	MIRROR RATIO: 1.090	
NEGATIVE IONS:	ON	ACQUISITION CAL FILE: DEFAULT	PSD MIRROR RATIO:

FIG. 6A

12/16



SUBSTITUTE SHEET (RULE 26)

COMMENT: 51305, PHOS I, 3', 40-MER(h-1)

METHOD: DNA1
MODE: LINEAR

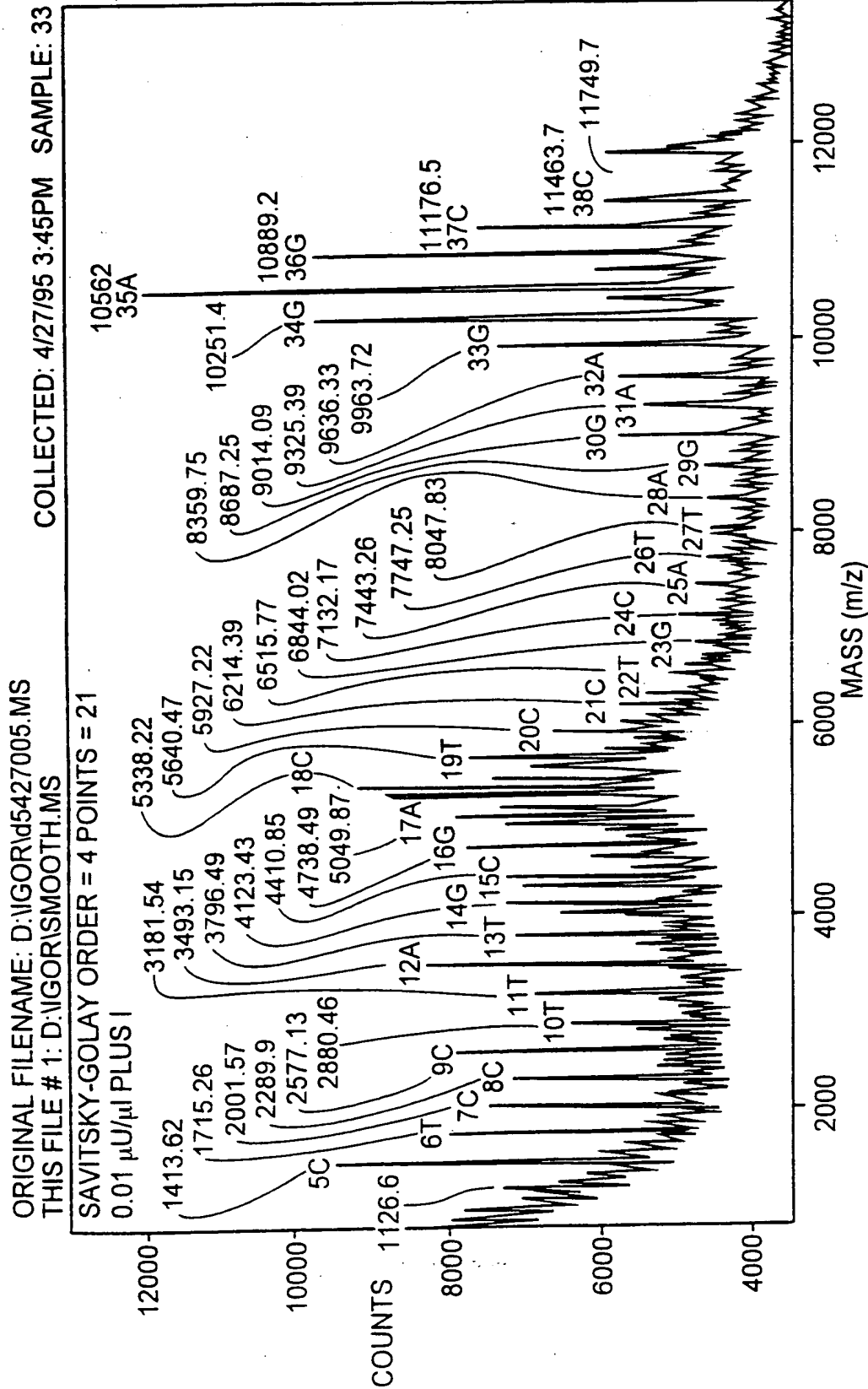
ACCELERATING VOLTAGE: 20000
GRID VOLTAGE: 93.0%
GUIDE WIRE VOLTAGE: 0.200%
NEGATIVE IONS: ON

LASER: 0
SCANS AVERAGED: 46
PRESSURE: 5.35e-07
ACQUISITION CAL FILE: D:\IGOR\40_CAL_USR

LOW MASS GATE: 800.0
TIMED ION SELECTOR:
MIRROR RATIO: 1.090
PSD MIRROR RATIO:

FIG. 6B

13/16



SUBSTITUTE SHEET (RULE 26)

COMMENT: PHOS I, 3', 40-MER(h-1)

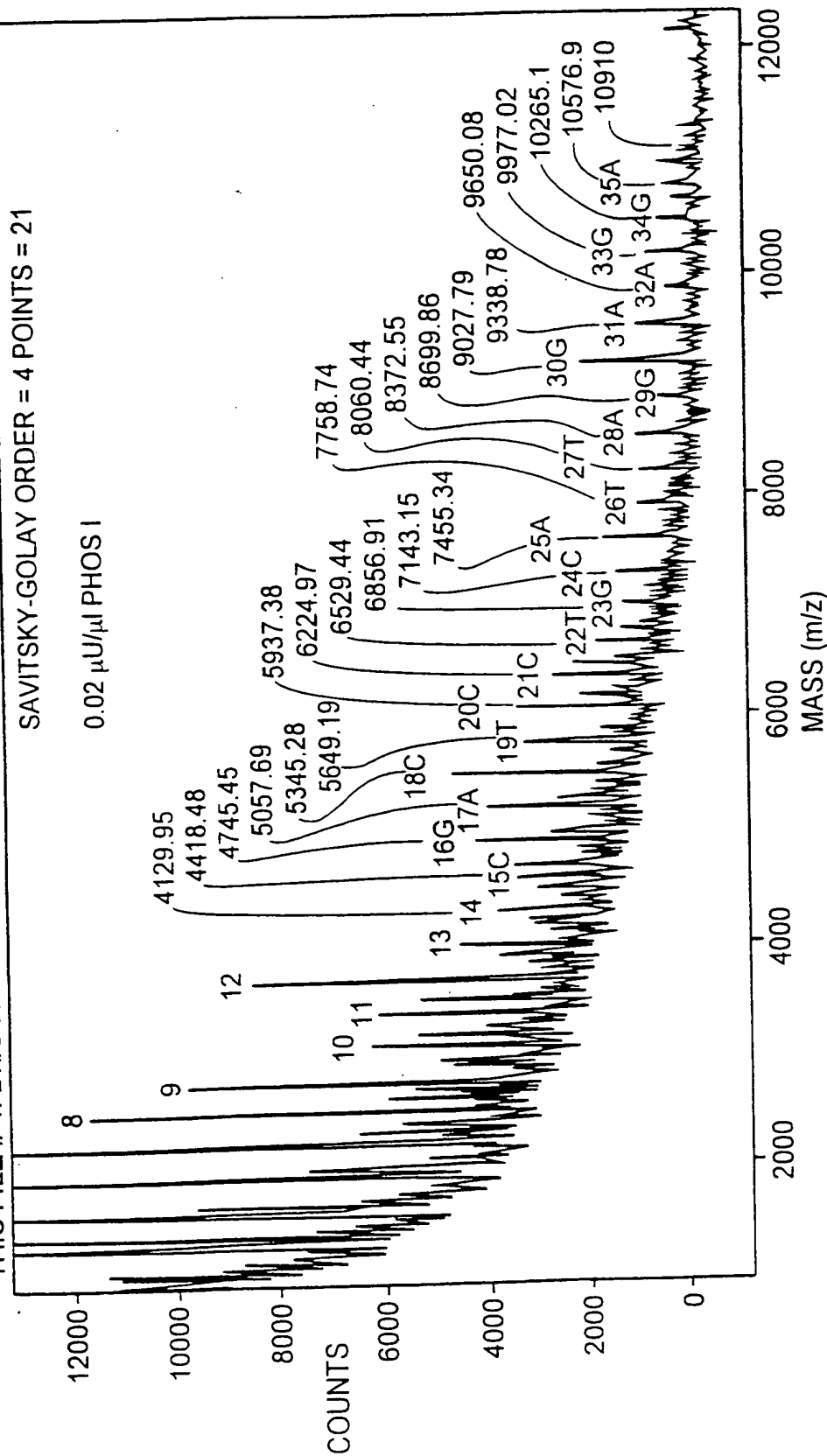
METHOD: DNA1 ACCELERATING VOLTAGE: 20000 LASER: 0 LOW MASS GATE: 800.0
MODE: LINEAR GRID VOLTAGE: 93.0% SCANS AVERAGED: 72 TIMED ION SELECTOR:
GUIDE WIRE VOLTAGE: 0.200% PRESSURE: 4.91e-07 MIRROR RATIO: 1.090
NEGATIVE IONS: ON ACQUISITION CAL FILE: D:\IGOR\40_CAL_USR PSD MIRROR RATIO:

FIG. 6C

14/16

ORIGINAL FILENAME: D:\IGOR\5427_6A.MS
THIS FILE # 1: D:\IGOR\SMOOTH.MS

COLLECTED: 4/27/95 4:13PM SAMPLE: 34



SUBSTITUTE SHEET (RULE 26)

COMMENT: 51320, PHOS I, 3', 40-MER(h-1)

METHOD: DNA1
MODE: LINEAR

ACCELERATING VOLTAGE: 20000
GRID VOLTAGE: 93.0%
GUIDE WIRE VOLTAGE: 0.200%
NEGATIVE IONS: ON

LASER: 0
SCANS AVERAGED: 58
PRESSURE: 4.02e-07

LOW MASS GATE: 800.0
TIMED ION SELECTOR:
MIRROR RATIO: 1.090
ACQUISITION CAL FILE: D:\IGOR\40_CAL_USR

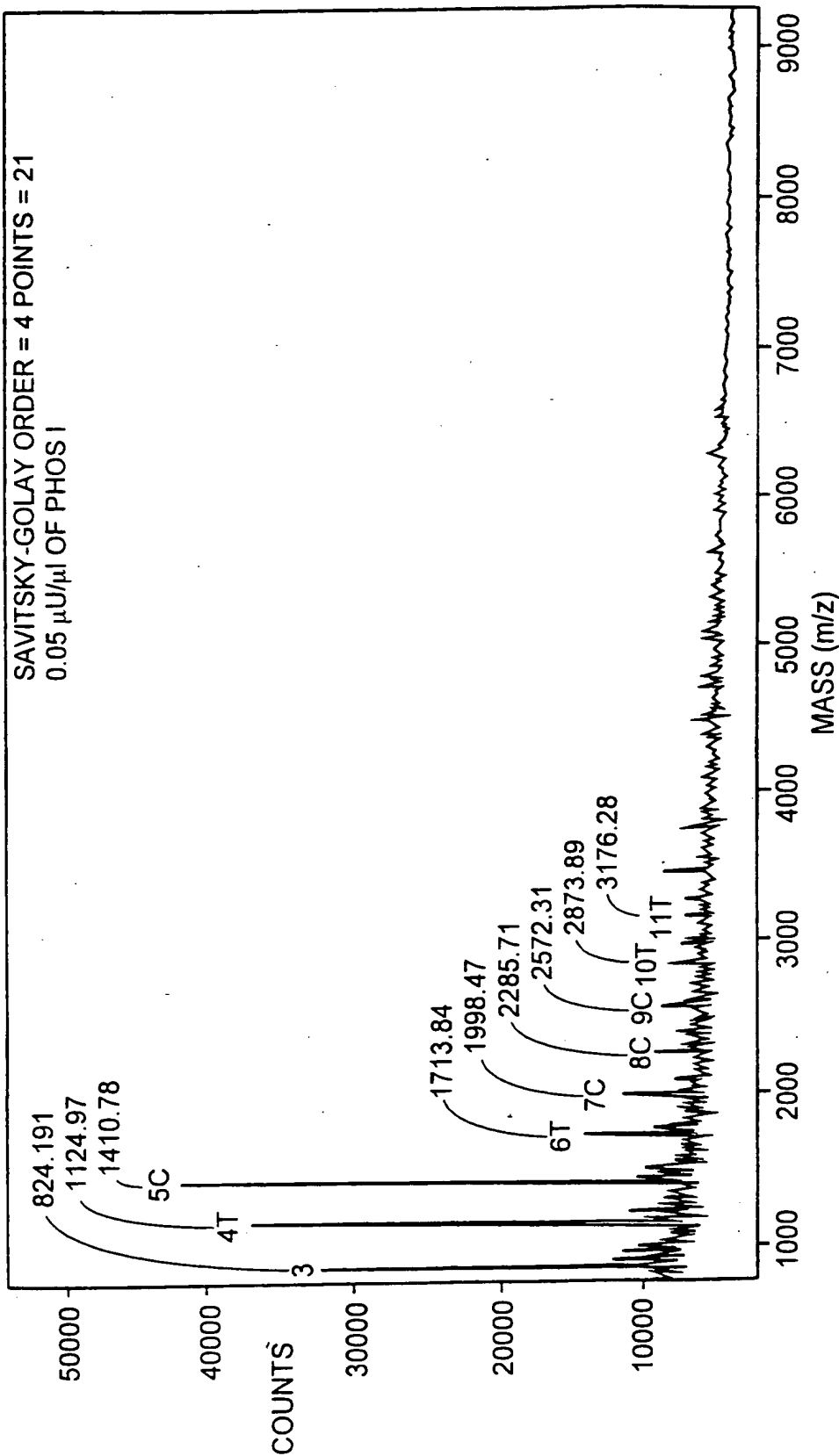
FIG. 6D

PSD MIRROR RATIO:

ORIGINAL FILENAME: D:\IGOR\ID5427008.MS
THIS FILE # 6: D:\IGOR\SMOOTH.MS

COLLECTED: 4/27/95 7:43PM SAMPLE: 36

SAVITSKY-GOLAY ORDER = 4 POINTS = 21
0.05 μ U/ μ l OF PHOS I



15/16

COMMENT: 51350, PHOS I, 3', 40-MER(h-1)

METHOD: DNA1 ACCELERATING VOLTAGE: 20000 LASER: 0
MODE: LINEAR GRID VOLTAGE: 93.0% SCANS AVERAGED: 30
GUIDE WIRE VOLTAGE: 0.200% PRESSURE: 3.66e-07
NEGATIVE IONS: ON ACQUISITION CAL FILE: DEFAULT

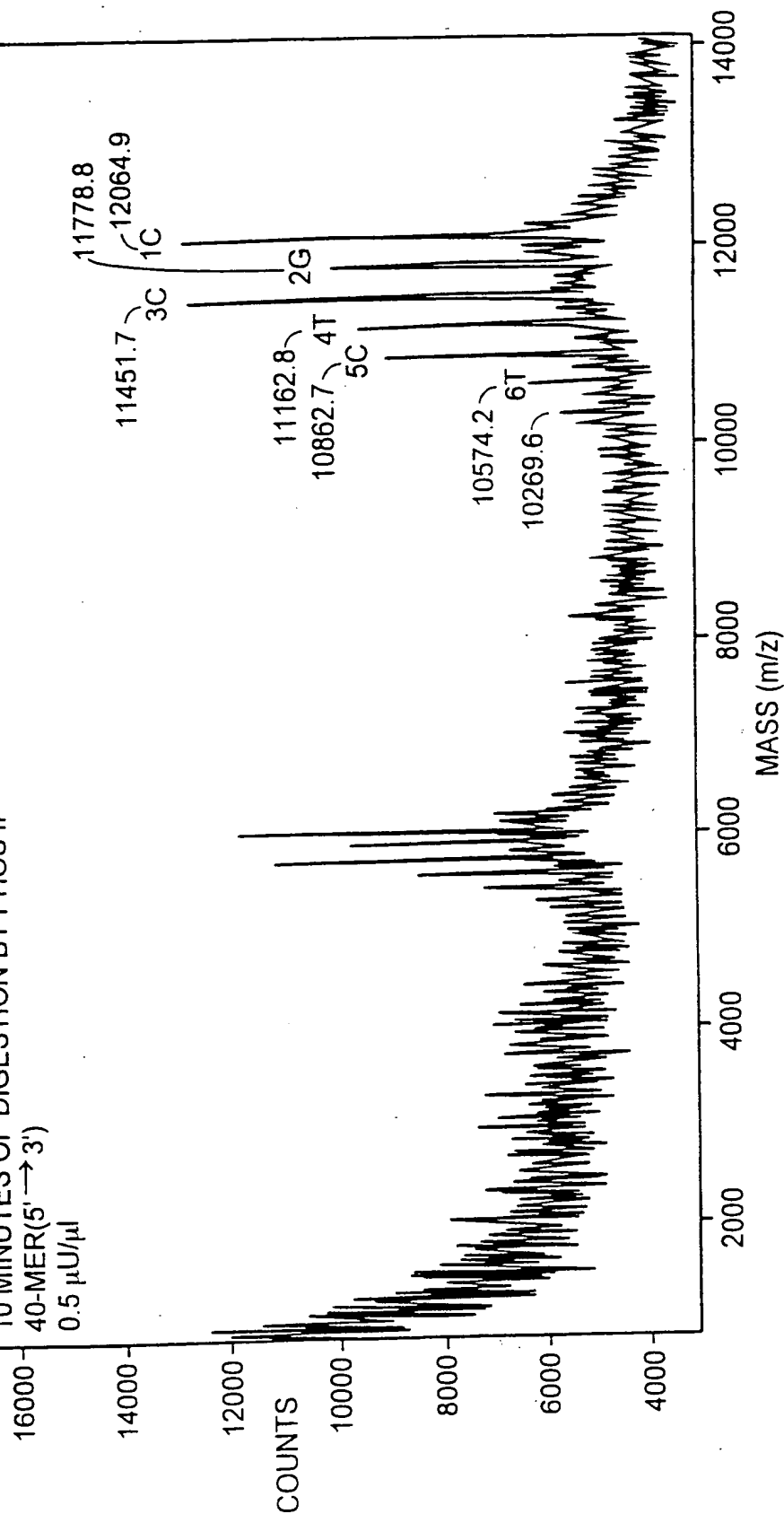
LOW MASS GATE: 800.0
TIMED ION SELECTOR:
DELAY: 100 ms
MIRROR RATIO: 1.090
PSD MIRROR RATIO:

FIG. 6E

ORIGINAL FILENAME: D:\IGORID5427012.MS
THIS FILE # 10: D:\IGORISMOOTH.MS

COLLECTED: 4/27/95 10:55PM SAMPLE: 68

SAVITSKY-GOLAY ORDER = 4 POINTS = 21
10 MINUTES OF DIGESTION BY PHOS II
40-MER(5' → 3')
0.5 μ U/ μ l



SUBSTITUTE SHEET (RULE 26)

COMMENT: 311, PHOS II, 10', 40-MER(h-1)

METHOD: DNA1	ACCELERATING VOLTAGE: 20000	LASER: 0	LOW MASS GATE: 800.0
MODE: LINEAR	GRID VOLTAGE: 93.0%	SCANS AVERAGED: 9	TIMED ION SELECTOR:
GUIDE WIRE VOLTAGE: 0.200%	PRESSURE: 4.96e-07	MIRROR RATIO: 1.090	MIRROR RATIO: 1.090
NEGATIVE IONS:	ON	ACQUISITION CAL FILE: DEFAULT	PSD MIRROR RATIO:

FIG. 7

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁶ : H01J 49/04, C12Q 1/00, G01N 33/48	A1	(11) International Publication Number: WO 96/36986 (43) International Publication Date: 21 November 1996 (21.11.96)
(21) International Application Number: PCT/US96/07146 (22) International Filing Date: 17 May 1996 (17.05.96) (30) Priority Data: 08/446,055 19 May 1995 (19.05.95) US 08/447,175 19 May 1995 (19.05.95) US (71) Applicant: PERSEPTIVE BIOSYSTEMS, INC. [US/US]; 500 Old Connecticut Path, Framingham, MA 01701 (US). (72) Inventors: PATTERSON, Dale, H.; Apartment J, 38 Bay Ridge Drive, Nashua, NH 03062 (US). TARR, George, E.; 640 Essex Street, S. Hamilton, MA 01982 (US). (74) Agent: TURANO, Thomas, A.; Testa, Hurwitz & Thibault, L.L.P., High Street Tower, 125 High Street, Boston, MA 02110 (US).		(81) Designated States: JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i> <i>With amended claims.</i> Date of publication of the amended claims: 9 January 1997 (09.01.97)
(54) Title: METHODS AND APPARATUS FOR SEQUENCING POLYMERS WITH A STATISTICAL CERTAINTY USING MASS SPECTROMETRY		
(57) Abstract The method and apparatus disclosed herein are useful for sequencing polymers using mass spectrometry. The methods involve differing ratios of hydrolyzing agent to polymer disposed upon a reaction surface adapted for use with a mass spectrometer. The methods further involve integrating data obtained from mass spectrometry analysis of a plurality of series of hydrolyzed polymer fragments, and provide statistical interpretation paradigms and computer software therefor. The apparatus involves a mass spectrometer sample holder, having hydrolyzing agent disposed thereon, which is useful for adapting any mass spectrometer for polymer sequencing.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

67

AMENDED CLAIMS

[received by the International Bureau on 3 December 1996 (03.12.96);
original claims 1, 3-7, 18, 19, 43, 45, 46 and 61 amended; new claims 108 and 109 added;
remaining claims unchanged (14 pages)]

1. A method of obtaining sequence information about a polymer comprising a plurality of monomers of known mass, said method comprising the steps of:
 - a) providing a set of polymer fragments, each differing by one or more monomers;
 - b) measuring a difference x between the mass-to-charge ratio of at least one pair of fragments;
 - c) asserting a mean difference μ between the mass-to-charge ratio of the pair of fragments measured in step b, wherein μ corresponds to a known mass-to-charge ratio of one or more differing monomers;
 - d) selecting a desired confidence level for μ ;
 - e) analyzing x to determine if it is statistically different from μ at the selected confidence level; and
 - f) determining if the asserted mean μ is assignable to the mass difference x with the selected confidence level based upon the analysis in step e).
2. The method of claim 1 wherein a statistical difference determined in the analysis of step e) indicates that the asserted mean μ is not assignable to the mass difference x with the selected confidence level.
3. The method of claim 1 comprising repeating steps c) through f) for a plurality of desired values of μ s.

4. The method of claim 1 wherein the analysis of step e) comprises a two-tailed t-test for an experimental mean.
5. The method of claim 1 wherein the analyzing in step e) comprises:
 - g) repeating step b) a number of times, n , to determine a measured mean mass-to-charge ratio difference \bar{x} between at least one pair of fragments;
 - h) determining a standard deviation s of the mean mass-to-charge ratio difference \bar{x} determined in step g);
 - i) comparing \bar{x} to the asserted mean difference μ ;
 - j) repeating steps c) through i) for a plurality of desired values of μ .

6. The method of claim 5 comprising repeating steps b) through j) for additional pairs of fragments.
7. The method of claim 5 wherein the comparing in step i) comprises taking the absolute value of the asserted mean difference.
8. The method of claim 5 further comprising the step of determining the number of measurements, n, based upon the analysis in step e).
9. The method of claim 1 wherein the polymer is a biopolymer.
10. The method of claim 9 wherein the biopolymer is selected from the group consisting of DNAs, RNAs, PNAs, proteins, peptides, carbohydrates and modified forms thereof.
11. The method of 1 further comprising the step of hydrolyzing the polymer to obtain the polymer fragments in step a).
12. The method of claim 1 further comprising hydrolyzing, on a reaction surface, the polymer with a hydrolyzing agent.
13. The method of claim 12 wherein the polymer is hydrolyzed on a reaction surface, said surface providing differing amounts of a hydrolyzing agent which hydrolyzes said polymer thereby to break inter-monomer bonds.
14. The method of claim 11, 12 or 13 wherein the hydrolyzing agent is an exohydrolase or an endohydrolase.
15. The method of claim 14 wherein hydrolyzing with said exohydrolase produces a series of fragments comprising a sequence-defining ladder of said polymer.
16. The method of claim 15 wherein the exohydrolase is selected from the group consisting of: exonucleases, exoglycosidases, and exopeptidases.

17. The method of claim 16 wherein the exopeptidase is selected from the group consisting of carboxypeptidase Y, carboxypeptidase A, carboxypeptidase B, carboxypeptidase P, aminopeptidase 1, leucine aminopeptidase, proline aminodipeptidase and cathepsin C.

AMENDED SHEET (ARTICLE 19)

18. The method of claim 16 wherein the exoglycosidase is selected from the group consisting of
- a) α - Mannosidase I
 - b) α - Mannosidase
 - c) β - Hexosaminodase
 - d) β - Galactosidase
 - e) α - Fucosidase I and II
 - f) α - Galactosidase
 - g) α - Neuraminidase and
 - h) α - Glucosidase I and II.
19. The method of claim 16 wherein the exonuclease is selected from the group consisting of
- a) λ - exonuclease
 - b) t7 Gene 1 exonuclease
 - c) exonuclease III
 - d) Exonuclease I
 - e) Exonuclease V
 - f) Exonuclease II and
 - g) DNA Polymerase II.
20. The method of claim 14 wherein hydrolyzing with said endohydrolase produces a series of fragments defining a map of said polymer.
21. The method of claim 20 wherein said endohydrolase is an endopeptidase selected from the group consisting of: trypsin, chymotrypsin, endo-proteinase Lys-C, endoproteinase Arg-C and thermolysin.

22. The method of claim 12 wherein the agent is a hydrolyzing agent other than an enzyme.
23. The method of claim 12 wherein said agent capable of hydrolyzing said polymer comprises a combination of at least one enzyme and at least one agent other than an enzyme.
24. The method of claim 13 wherein the reaction surface comprises an array of discrete separable zones, each zone comprising a differing amount of said hydrolyzing agent.

38. The method of claim 12 wherein hydrolysis is accomplished by immobilizing said agent on said reaction surface.
39. The method of claim 12 wherein hydrolysis is accomplished using a hydrolyzing agent in liquid or gel form, said liquid or gel form being resistant to physical dislocation.
40. The method of claim 1 comprising the additional step of combining a light-absorbent matrix with said fragments prior to step b).
41. The method of claim 1 comprising the additional step of combining said polymer fragments with moieties for selectively shifting the mass of hydrolyzed sequences prior to step b).
42. The method of claim 1 comprising the additional step of combining said polymer fragments with moieties for improving ionization prior to step b).
43. A method for obtaining sequence information about a polymer comprising a series of different monomers of known mass, said method comprising the steps of:
 - a) providing a set of polymer fragments, each differing by one or more monomers;
 - b) measuring the mass-to-charge ratio difference x between a pair of fragments;
 - c) asserting a mean difference μ , which is related to a known mass-to-charge ratio of one or more monomers;
 - d) selecting a desired confidence level for μ ;

- e) repeating step b) to obtain a number of measurements n , thereby to determine the measured mean mass-to-charge ratio difference \bar{x} between the pair of fragments;
- f) determining the standard deviation s of the measured mean mass-to-charge ratio difference \bar{x} determined in step e;
- g) calculating a test statistic $t_{\text{calculated}}$ with the following algorithm:

$$t_{\text{calculated}} = \frac{|\bar{x} - \mu| \sqrt{n}}{s};$$

- h) comparing the test statistic $t_{\text{calculated}}$ calculated in step g to a t-distribution corresponding to the number of measurements and the desired confidence level; and
- i) determining if the asserted mean μ is assignable to the mass difference x with the selected confidence level based upon the comparison in step h.

AMENDED SHEET (ARTICLE 19)

44. The method of claim 43 further comprising a comparison of the calculated test statistic $t_{\text{calculated}}$ in step g) to a t-distribution corresponding to the number of measurements and the desired confidence level.
45. The method of claim 43 further comprising repeating steps b) - i) for additional pairs of fragments thereby to obtain sequence information.
46. The method of claim 43 further comprising the step of determining the number of measurements, n , based upon the comparison in step h).
47. The method of claim 43 wherein the polymer is a biopolymer.
48. The method of claim 47 wherein the biopolymer is selected from the group consisting of DNAs, RNAs, PNAs, proteins, peptides, carbohydrates and modified forms thereof.
49. The method of claim 43 further comprising the step of hydrolyzing the polymer with a hydrolyzing agent to create the fragments in step a).
50. The method of claim 49 wherein the hydrolyzing agent is an exohydrolase which produces a series of fragments comprising a sequence-defining ladder of said polymer.
51. The method of claim 50 wherein the exohydrolase is selected from the group consisting of: exonucleases, exoglycosidases, exopeptidases.
52. The method of claim 51 wherein the exopeptidase is selected from the group consisting of carboxypeptidase Y, carboxypeptidase A, carboxypeptidase B, carboxypeptidase P, aminopeptidase 1, leucine aminopeptidase, proline aminodipeptidase and cathepsin C.

53. The method of claim 51 wherein the exoglycosidase is selected from the group consisting of

- a) α - Mannosidase I
- b) α - Mannosidase
- c) β - Hexosaminidase
- d) β - Galactosidase
- e) α - Fucosidase I and II
- f) α - Galactosidase

AMENDED SHEET (ARTICLE 19)

- g) α - Neuraminidase and
 - h) α - Glucosidase I and II.
54. The method of claim 51 wherein the exonuclease is selected from the group consisting of
- a) Exonuclease
 - b) λ - exonuclease
 - c) t7 Gene 1 exonuclease
 - d) exonuclease III
 - e) Exonuclease I
 - f) Exonuclease V
 - g) Exonuclease II
 - h) DNA Polymerase II.
55. The method of claim 49 wherein the hydrolyzing agent is other than an enzyme.
56. The method of claim 49 wherein the agent comprises a combination of at least one enzyme and at least one agent other than an enzyme.
57. The method of claim 49 wherein hydrolysis is performed on a reaction surface, said surface providing differing amounts of a hydrolyzing agent.
58. The method of claim 57 wherein the reaction surface comprises an array of discrete separable zones, each zone comprising a differing amount of said hydrolyzing agent.
59. The method of claim 49 wherein the reaction surface comprises a continuous concentration gradient of a hydrolyzing agent.
60. The method of claim 43 further comprising adding a matrix to the polymer fragments before measuring the mass-to-charge ratio in step b).

61. A method for obtaining sequence information about a polymer having a plurality of monomers of known mass, said method comprising:
- providing a set of polymer fragments, each differing by one or more monomers;
 - measuring a difference x between the mass-to-charge ratio of a pair of fragments;
 - asserting a mean difference μ between the mass-to-charge ratio of the pair of fragments measured in step b, wherein μ corresponds to a known mass-to-charge ratio of one or more monomers;
 - selecting the desired confidence level for μ ;
 - analyzing x to determine if it is statistically different from μ at the selected confidence level;
 - repeating steps b)-e) a number of times n , until a plurality of desired values of μ s have been asserted;
 - determining if the asserted mean μ is assignable to the mass difference x with the selected confidence level based upon the analysis in step e; and
 - repeating steps b) -g) for additional pairs of fragments.
62. The method of claim 61 wherein the polymer is a biopolymer.
63. The method of claim 62 wherein the biopolymer is selected from the group consisting of DNAs, RNAs, PNAs, proteins, peptides, carbohydrates and modified forms thereof.
64. The method of claim 61 wherein the polymer fragments in step a) are created by concentration dependent hydrolysis of the polymer.
65. The method of claim 61 further comprising the step of hydrolyzing said polymer with a hydrolyzing agent to produce the polymer fragments in step a).

AMENDED SHEET (ARTICLE 19)

66. The method of claim 65 wherein the hydrolyzing agent is an exohydrolase.
67. The method of claim 66 wherein the hydrolysis caused by said exohydrolase produces a series of fragments defining a ladder of said polymer.
68. The method of claim 66 wherein the exohydrolase is selected from the group consisting of: exonucleases, exoglycosidases, and exopeptidases.
69. The method of claim 68 wherein the exoglycosidase is selected from the group consisting of
 - a) α - Mannosidase I
 - b) α - Mannosidase
 - c) β - Hexosaminidase
 - d) β - Galactosidase
 - e) α - Fucosidase I and II
 - f) α - Galactosidase
 - g) α - Neuraminidase and
 - h) α - Glucosidase I and II.

107. The sample plate of any one of claims 78, 85, 99, 100 or 101 wherein said plate is disposable.
108. A method of obtaining information about the identity of a polymer comprising a plurality of monomers of known mass, said method comprising the steps of:
- a) providing a set of polymer fragments created by the endohydrolysis of said polymer;
 - b) measuring the mass-to-charge ratio of a fragment;
 - c) asserting a hypothetical identity for the fragment, wherein the hypothetical identity corresponds to a known identity of a fragment of a reference polymer, said fragment having a known mass-to-charge ratio;
 - d) selecting a desired confidence level for the hypothetical identity; and
 - e) determining whether the measured mass-to-charge ratio is statistically different from the mass-to-charge ratio of the asserted hypothetical fragment;
 - f) determining if the asserted hypothetical identity is assignable to the measured mass-to-charge ratio of the fragment with the selected confidence level based upon the determination in step e; and
 - g) repeating steps b)- e).
109. The method of claim 108 wherein the hypothetical identity in step c) corresponds to a known identity derived from a computer database of known sequences.